Nod-factor perception

This application is the National Stage application of PCT/DK2004/000478 filed on July 2, 2004, which claims benefit to Danish Application No. PA 2003

5   01010 filed on July 3, 2003, which application(s) are incorporated herein by reference.

Field of the invention

The invention relates to a novel Nod-factor binding element and component

10   polypeptides that are useful in enhancing Nod-factor binding in nodulating plants and inducing nodulation in non-nodulating plants. More specifically, the invention relates to Nod-factor binding proteins and their respective genomic and mRNA nucleic acid sequences.

15   **Background of the invention**

The growth of agricultural crops is almost always limited by the availability of nitrogen, and at least 50% of global needs are met by the application of synthetic fertilisers in the form of ammonia, nitrate or urea. Apart from recycling of crop residues and animal manure, and atmospheric deposition,

20   the other most important source of nitrogen for agriculture comes from biological nitrogen fixation.

A small percentage of prokaryots, the diazotrophs, produce nitrogenases and are capable of nitrogen fixation. Members of this group, belonging to the *Rhizobiaceae* family (for example *Mesorhizobium loti*, *Rhizobium meliloti*,

25   *Bradyrhizobium japonicum*, *Rhizobium leguminosarum* bv *viceae*) here collectively called *Rhizobium* or *Rhizobia* spp and the actinobacterium *Frankia* spp, can form endosymbiotic associations with plants conferring the ability to fix nitrogen. Although many plants can associate with nitrogen fixing bacteria, only a few plants, all members of the Rosid I Clade, form

30   endosymbiotic associations with *Rhizobia* spp and *Frankia* spp., which are unique in that most of the nitrogen is transferred to and assimilated by the

host plant. Legumes, including soybean, bean, pea, peanut, chickpea, cowpea, lentil, pigeonpea, alfalfa and clover, are the most agronomically important members of this small group of nitrogen-fixing plants.

The rhizobial-legume interaction is generally host-strain specific, whereby

5   successful symbiotic associations only occur between specific rhizobial strains and a limited number of legume species. The specificity of this interaction is determined by chemical signalling between plant and bacteria, which accompanies the initial interaction and the establishment of the symbiotic association (Hirsch *et al.* 2001, *Plant Physiol.* 127: 1484-1492).

10   Specific (iso)flavanoids, secreted into the soil by legume spp, allow *Rhizobium* spp to distinguish compatible hosts in their proximity and to migrate and associate with roots of the host. In a compatible interaction, the (iso)flavanoid perceived by the *Rhizobium* spp, interacts with the rhizobial *nodD* gene product, which in turn leads to the induction of rhizobial Nod-

15   factor synthesis. Nod-factor molecules are lipo-chitin-oligosaccharides, commonly comprising four or five β-1-4 linked N-acetylglucosamines, with a 16 to 18 carbon chain fatty acid n-acetylated on the terminal non-reducing sugar. Nod factors are synthesised in a number of variants, characterised by their chemically different substitutions on the chitin backbone which are

20   distinguished by the compatible host plant. The perception of Nod-factors by the host induces invasion zone root hairs, in the proximity of rhizobial cells, to curl and entrap the bacteria. The adjacent region of the root hair plasma membrane invaginates and new cell wall material is synthesized to form an infection thread or tube, which serves to transport the symbiotic bacteria

25   through the epidermis to the cortical cells of the root. Here the cortical cells are induced to divide to form a primordium, from which a root nodule subsequently develops. In legumes belonging to genera like *Arachis* (peanut), *Stylosantos* and *Sesbania,* infection is initiated by a simple "crack entry" through spaces or cavities between epidermal cells and lateral roots.

30   In spite of these differences, perception of Nod factors by the host plant simultaneously induces the expression of a series of plant nodulin genes,

which control the development and function of root nodules, wherein the rhizobial endosymbiotic association and nitrogen fixation are localised.

A variety of molecular approaches have identified a series of plant nodulin genes which play a role in rhizobial-legume symbiosis, and whose

5   expression is induced at early or later stages of rhizobial infection and nodule development (Geurts and Bisseling, 2002, *Plant Cell* supplement S239-249). Furthermore, plant mutant studies have revealed that a signalling pathway must be involved in amplifying and transducing the signal resulting from nod-factor perception, which is required for the induction of nodulin gene

10   expression. Among the first physiological events identified in this signal transduction pathway, which occurs circa 1 min after Nod-factor application to the root epidermis, is a rapid calcium influx followed by chloride efflux, causing depolarisation of the plasma membrane and alkalization of the external root hair space of the invasion zone. A subsequent efflux of

15   potassium ions allows re-polarisation of the membrane, and later a series of calcium oscillations are seen to propagate the signal through the root hair cell. Pharmacological studies with specific drugs, which mimic or block Nod-factor induced responses, have identified potential components of the signalling pathway. Thus mastoparan, a peptide which is thought to mimic

20   the activated intracellular domain of G-protein coupled receptors, can induce early Nod gene expression and root hair curling. This suggests that trimeric G protein may be involved in the Nod-factor signal transduction pathway. Analysis of a group of nodulation mutants, including some that fail to show calcium oscillations in response to Nod-factor signals, has revealed that in

25   addition to the lack of nodulation, these mutants are unable to form endosymbioses with arbuscular mycorrhizal fungi. This implies that a common symbiotic signal transduction pathway is shared by two types of endosymbiotic relationships, namely root nodule symbiosis, which is largely restricted to the legume family, and arbuscular mycorrhizal symbiosis, which

30   is common to the majority of land plant species. This suggests that there may

be a few key genes which dispose legumes to engage in nodulation, and which are missing from crop plants such as cereals.

The identification of these key genes, which encode functions which are indispensable for establishing a nitrogen fixing system in legumes, and their transfer and expression in non-nodulating plants, has long been a goal of molecular plant breeders. This could have a significant agronomic impact on the cultivation of cereals such as rice, where production of two harvests a year may require fertilisation with up to 400 kg nitrogen per hectare. In accordance with this goal, WO02102841 describes the gene encoding the NORK polypeptide, isolated from the nodulating legume *Medicago sativa*, and the transformation of this gene into plants incapable of nitrogen fixation. The NORK polypeptide and its homologue/orthologue SYMRK from *Lotus japonicus* (Stracke *et al* 2002 *Nature* 417:959-962), are transmembrane receptor-like kinases with an extracellular domain comprising leucine-rich repeats, and an intracellular protein kinase domain. *Lotus japonicus* mutants, with a non-functional *SYMRK* gene, fail to form symbiotic relationships with either nodulating rhizobia or arbuscular mycorrhiza. This implies that a common symbiotic signalling pathway mediates these two symbiotic relationships, where SYMRK comprises an early step in the pathway. The *symRK* mutants retain an initial response to rhizobial infection, whereby the root hairs in the susceptable invasion zone undergo swelling of the root hair tip and branching, but fail to curl. This suggests that the SYMRK protein is required for an early step in the common symbiotic signalling pathway, located downstream of the perception and binding of microbial signal molecules (e.g. Nod-factors), that leads to the activation of nodulin gene expression.

The search for key symbiosis genes has also focussed on 'candidate genes' encoding receptor proteins with the potential for perceiving and binding Nod-factors or surface structures on rhizobial bacteria. US 6,465,716 discloses NBP46, a Nod-factor binding lectin isolated from *Dolichos biflorus* roots, and its transgenic expression in transformed plants. Transgenic expression of

NBP46 in plants is reported to confer the ability to bind to specific carbohydrates in the rhizobial cell wall and thereby to bind these bacteria and utilise atmospheric nitrogen, as well as conferring apyrase activity. An alternative approach to search for key symbiosis genes has been to screen

5   for Nod-factor binding proteins in protein extracts of plant roots. NFBS1 and NFBS2 were isolated from *Medicago trunculata* and shown to bind Nod-factors in nanomolar concentrations, however, they both failed to exhibit the Nod-factor specificity characteristic of rhizobial-legume interactions (Geurts and Bisseling, 2002 *supra*).

10

The Nod-factor binding element, which is responsible for strain specific Nod-factor perception is not, as yet, identified. The isolation and characterisation of this element and its respective gene(s) would open the way to introducing Nod-factor recognition into non-nodulating plants and thereby the potential to

15   establish *Rhizobium*-based nitrogen fixation in important crop plants.

Rhizobial strains produce strain-specific Nod-factors, lipochitin oligosaccharides (LCOs), which are required for a host-specific interaction with their respective legume hosts. *Lotus* and peas belong to two different

20   cross-inoculation groups, where *Lotus* develops nodules after infection with *Mesorhizobium loti,* while pea develops nodules with *Rhizobium leguminosarum* bv *viceae*. Cultivars belonging to a given *Lotus* sp also vary in their ability to interact and form nodules with a given rhizobial strain. Perception of Nod-factor secreted by *Rhizobium* spp bacteria, as the first

25   step in nodulation, commonly leads to the initiation of tens or even hundreds of rhizobial infection sites in a root. However, the majority of these infections abort and only in a few cases do the rhizobia infect the nodule primordium. The frequency and efficiency of the *Rhizobium*-legume interaction leading to infection is known to be influenced by variations in Nod-factor structure. The

30   genetics of Nod-factor synthesis and modification of their chemical structure in *Rhizobium* spp have been extensively characterised. An understanding of

Nod-factor binding and perception, and the structure of its component elements is needed in order to optimise the host Nod-factor response. This information would, in turn, provide the necessary tools to breed for enhanced efficiency of nodulation and nitrogen fixation in current nitrogen-fixing crops.

5

The importance of this goal is clearly illustrated by the performance of the major US legume crop, soybean, which is grown on 15%, or more, of agricultural land in the US. While nitrogen fixation by soybean root nodules can assimilate as much as 100 kg nitrogen per hectare per year, these high

10   levels of nitrogen assimilation are insufficient to support the growth of the highest yielding modern soybean cultivars, which still require the application of fertiliser.

In summary, there is a need to increase the efficiency of nodulation and nitrogen fixation in current legume crops as well as to transfer this ability to

15   non-nodulating crops in order to meet the nutritional needs of a growing global population, while minimising the future use of nitrogen fertilisers and their associated negative environmental impact.

Summary of the invention

20   The invention provides an isolated Nod-factor binding element comprising one or more isolated NFR polypeptide having a specific Nod-factor binding property, or a functional fragment thereof, wherein the NFR amino acid sequence is at least 60% identical to either of SEQ ID NO: 8, 15 or 25. The isolated NFR polypeptides of the invention include NFR1, comprising an

25   amino acid sequence selected from the group consisting of SEQ ID No: 24, 25, 52 and 54, having specific Nod-factor binding properties, and NFR5 comprising an amino acid sequence selected from the group consisting of SEQ ID No: 8, 15, 32, 40 and 48, having specific Nod-factor binding properties. Furthermore, the invention provides an isolated nucleic acid

30   molecule encoding a NFR1 polypeptide or a NFR5 polypeptide of the invention, and an expression cassette, and vector and transformed cell

comprising said isolated nucleic acid molecule. In a further embodiment is provided a nucleic acid molecule encoding a NFR polypeptide of the invention that hybridises with a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID No: 6, 7, 11, 12, 21,

5   22, 23, 39, 47, 51 and 53.

According to a further embodiment of the invention, a method is provided for producing a plant expressing the Nod-factor binding element, the method comprising introducing into the plant a transgenic expression cassette comprising a nucleic acid sequence, encoding the NFR polypeptide of the

10   invention, wherein the nucleic acid sequence is operably linked to its own promoter or a heterologous promoter, preferably a root specific promoter. In a preferred embodiment, the expression of both said NFR 1 and NFR5 polypeptides by the transgenic plant confers on the plant the ability to bind Nod-factors in a chemically specific manner and thereby initiate the

15   establishment of a *Rhizobium*-plant interaction leading to the development of nitrogen-fixing root nodules.

According to a further embodiment, the invention provides a method for marker assisted breeding of *NFR* alleles, encoding variant NFR polypeptides, comprising the steps of identifying variant NFR polypeptides in a nodulating

20   legume species, comprising an amino acid sequence substantially similar to variant NFR polypeptide having specific Nod-factor binding properties and having an amino acid sequence selected from the group consisting of SEQ ID No: 8, 15, 24, 25, 32, 40, 48, 51 and 53; determining the nodulation frequency of plants expressing said variant NRF polypeptide; identifying DNA

25   polymorphisms at loci genetically linked to or within the allele locus encoding said variant NFR locus; preparing molecular markers based on said DNA polymorphisms; and using said molecular markers for the identification and selection of plants carrying *NFR* alleles encoding said variant NFR polypeptides. The invention includes plants selected by the use of this

30   method of marker assisted breeding. In a preferred embodiment, said

method of marker assisted breeding of *NFR* alleles provides for the breeding legumes with enhanced nodulation frequency and nodule occupancy.

Brief Description of the figures

5 **Figure 1:** Map based cloning of *Lotus NFR5*. **a**. Genetic map of the *NFR5* region with positions of linked AFLP and microsatellite markers above the line and distances in cM below. The fraction of recombinant plants detected in the mapping population is indicated. **b**. Physical map of the BAC and TAC clones between the closest linked microsatellite markers. The positions of

10 sequence-derived markers used to fine-map the *NFR5* locus, and the fraction of recombinant plants found in the mapping population are indicated. **c**. Candidate genes identified in the sequenced region delimited by the closest linked recombination events. **d**. Structure of the *NFR5* gene, position of the transcription initiation point and the *nfr5-1*, *nfr5-2* and *nfr5-3* mutations. The

15 asterisk indicates the position of a stop codon in *nfr5-3*; the black triangle a retrotransposon insertion in *nfr5-2*; and the grey box defines the deletion in *nfr5-1*. GGDP: geranylgeranyl diphosphate synthase; RE: retroelement; RZF: ring zinc finger protein; GT: glycosyl transferase; A2L: apetala2-like protein; RLK: receptor-like kinase; PL: pectate lyase-like protein; AS: ATPase-

20 subunit; HD: homeodomain protein; RF: ring finger protein. Hypothetical proteins are not labelled. **e**. Southern hybridization demonstrating deletion of SYM10 in the "N15" *sym10* mutant line. *Eco*RI digested genomic DNA of the parental variety "Sparkle" and the fast neutron derived mutant "N15" hybridized with a pea *SYM10* probe covering the region encoding the

25 predicted extracellular domain. Hybridization with a probe from the 3′untranslated region demonstrated that the complete gene was deleted.

Figure 2: Structure and domains of the NFR5 protein. **a**. Schematic representation of the NFR5 protein domains. **b**. The amino acid sequence of NFR5 arranged in protein domains. Bold, conserved LysM residues. Bold

30 and underlined residues conserved in protein kinase domains (KD); TM:

transmembrane, SP: signal peptide. The asterisk indicates a stop codon in the *nfr5-3*; the black triangle a retrotransposon insertion in *nfr5-2* and the grey box defines the amino acids deleted in *nfr5-1*. **c.** Individual alignment of the three LysM motifs (M1, M2, M3) from NFR5, pea SYM10, *Medicago*

5 *truncatula* (*M.t*, Ac126779) rice (Ac103891), the single LysM in chitinase from *Volvox carteri* (Acc. No: T08150) and the pfam consensus. **d.** The divergent or absent activation loop (domain VIII) in the NFR5 family of receptor kinases is illustrated by alignment of kinase motifs VII, VIII and IX from *Arabidopsis* (At2g33580) NFR5, SYM10, *Medicago truncatula* (*M.t*, Ac126779), rice

10 (Ac103891) and the SMART concensus. Conserved domain VII aspartic acid is marked in bold and underlined. **c** and **d** the amino acids conserved in all aligned motifs are marked in black and amino acids conserved in two or more motifs are marked in grey.

Figure 3. The aligned amino acid sequence of the LjNFR5 and PsSYM10

15 proteins. Amino acid residues sharing identity are highlighted. The *Medicago truncatula* (Ac126779) showing 76 % amino acid identy to *Lotus* NFR5 is included to exemplify a substantial identical protein sequence.

Figure 4. Steady-state levels of *LjNFR5* and *PsSYM10* mRNA. **a.** *NFR5* mRNA detected in uninoculated roots, inoculated roots, nodules, leaves,

20 flowers and pods of *Lotus* plants. **b.** Time course of *NFR5* mRNA transcript accumulation in roots after inoculation with *M. loti*. The identity of the amplified transcripts was confirmed by sequencing. ATPase was used as internal control and relative normalised values compared to uninoculated roots are shown. **c.** Northern analysis showing *NFR5* mRNA expression in

25 nodule leaf and root of symbiotically and non-symbiotically grown *Lotus* plants. **d.** Northern analysis showing *Sym10* mRNA expression in leaf, root and nodule of symbiotically and non-symbiotically grown pea plants.

Figure 5. Positional cloning of the *NFR1* gene. a. Genetic map of the region surrounding the *NFR1* locus. Positions of the closest AFLP, microsatelitte-

and PCR-markers are given together with genetic distances in cM. b. Physical map of the *NFR1* locus. BAC clones 56L2, 16K18, 10M24, 36D15, 56K22 and TAC clones LjT05B16, LjT02D13, LjT211O02, which cover the region are shown. The numbers of recombination events detected with BAC

5   and TAC end-markers or internal markers are given. Arrows indicate the positions of the two markers (10M24-2, 56L2-2) delimiting the *NFR1* locus. UFD and HP correspond to the UFD1-like protein and the hypothetical protein encoded in the region. c. Exon-intron structure of the *NFR1* gene. Boxes correspond to exons, where LysM motifs are shown in light grey,

10   trans-membrane region in black, kinase domains in dark grey. Dotted lines define introns and full lines define the 5' and 3' un-translated regions. The nucleotide length of all exons and introns are indicated. The numbers between brackets correspond to exon and intron 4, corresponding to alternative splicing.

15   **Figure 6.** Structure and domains of the NFR1 protein. **a.** Primary structure of the *NFR1* protein comprising a signal peptide (SP); LysM motifs (LysM1 and LysM2); transmembrane region (TM); protein kinase domains with conserved amino acids in bold and underlined (PK). The cysteine couples (**CxC**) are in bold and the LysM amino acids important for secondary structure

20   maintenance are underlined. The two extra amino acids resulting from alternative splicing are shown in brackets. I-XI represent the kinase domains. Asterisks indicate positions of the nonsense mutations found in *NFR1-1* and *NFR1-2* mutant alleles. **b.** Alignments of the two *NFR1* LysM motifs to the consensus sequences predicted by the SMART program and the

25   *Arabidopsis thaliana* (Acc No: NP566689), rice (*O. sativa*) (Acc No: BAB89226), and *Volvox carteri* (Acc. No: T08150) LysM motifs.

Figure 7. *NFR1*, *NFR5* and *SymRK* gene expression. **a.** Transcript levels of *NFR1* in uninoculated, inoculated roots, nodules, leaves, flowers and pods of

30   wild type plants. **b.** Transcript levels of *NFR1* in wild type, *nfr1, nfr5* and

symRK mutant plants after inoculation with *M. loti*. **c**. Transcript levels of *NFR5* in wild type, *nfr1, nfr5* and *SymRK* mutant plants after inoculation with *M. loti*. **d**. Transcript levels of *SYMRK* in wild type, *nfr1, nfr5* and *symRK* mutant plants after inoculation with *M. loti*. Transcript levels were measured

5   by quantitative PCR. ATPase was used as internal control and relative values normalised to the untreated roots (zero hours) are shown.

Figure 8. Root hair response after inoculation with *M. loti* or Nod-factor application. **a**. Wild type root hair curling on seedlings inoculated with *M. loti*.

10   **b**. Root hair deformations on wild type seedlings after Nod-factor application. **c**. Root hairs on *nfr1-1* seedlings inoculated with *M. loti*. **d**. Root hairs on *nfr1-1* seedlings after Nod-factor application. **e**. Root hairs with balloon deformations on *symRK-3* mutants inoculated with *M. loti*. **f**. Roots hairs on a *nfr1-1,symRK-3* double mutant inoculated with *M. loti* **g**. Excessive root hair

15   response on *nin* mutants inoculated with *M. loti*. **h**. Root hairs on a *nfr1-1,nin* double mutant inoculated with *M. loti*. Root hairs on *nfr5-1* seedlings inoculated with *M. loti, nfr5-1* seedlings after Nod-factor application, untreated *nfr5-1* control, untreated wild type control, untreated *nfr1-1* control, are indistiguisable from the straight roots hairs shown in **c, d, f, h** and

20   therefore not shown. Inserts to the right of **a** to **h** show a close-up of the root hairs.

Figure 9. Membrane depolarisation and pH changes in the extracellular root hair space after application of Nod-factor purified from *M. loti*. Influence of 0.1

25   µM Nod-factor (NF) on membrane potential (Em) and/or external pH (pH) of **a**. *Lotus* wild type **b**. *nfr5-1* and *nfr5-2* mutants **c**. *nfr1-1* and *nfr1-2* mutants **d**. *symRK-1* and *symRK-3* mutants **e**. *nfr1-2,symRK-3* double mutant, **f**. pH changes in the extracellular root hair space after application of an undecorated chito-octaose.

30

Figure 10. Expression of the *NIN* and *ENOD2* genes in wild type, *nfr1* and *nfr5* mutant genotypes. **a.** NIN transcript level in RNA extracted from roots two hours to 12 days after *M. loti* inoculation. **b** ENOD2 transcript level in RNA extracted from roots two hours to 12 days after *M. loti* inoculation.

5   Transcript levels were measured by quantitative PCR and the identity of the amplified sequences was confirmed by sequencing. ATPase was used as internal control and relative values normalised to the untreated root (zero hours) are shown.

10   **Figure 11.** Alignment NFR1 and NFR5 proteins reveal an overall similarity of 33 % amino acid identities

Figure 12. Domain structure of native and hybrid NFR1 and NFR5 polypeptides.

15   **Detailed description of the invention**

I. Definitions

AFLP: Amplified Fragment Length Polymorphism is a PCR-based technique for the amplification of genomic fragments obtained after digestion with two

20   different enzymes. Different genotypes can be differentiated based on the size of amplified fragments or by the presence or absence of a specific fragment (Vos, P. (1998), *Methods Mol Biol.*, 82:147-155). Amplified Fragment Length Polymorphism is a PCR-based technique used to map genetic loci.

25   ***Agrobacterium rhizogenes*-mediated transformation**: is a technique used to obtain transformed roots by infection with *Agrobacterium rhizogenes*. During the transformation process the bacteria transfers a DNA fragment (T-DNA) from an endogenous plasmid into the plant genome (Stougaard, J. *et al*, (1987) *Mol.Gen.Genet.* 207, 251-255). For transfer of a gene of interest

the gene is first inserted into the T-DNA region of *Agrobacterium rhizogenes* which is subsequently used for wound-site infection.

Allele: gene variant

BAC clones: clones from a Bacterial Artificial Chromosome library

5 **Conservatively modified variant**: when referring to a polypeptide sequence when compared to a second sequence, includes individual conservative amino acid substitutions as well as individual deletions, or additions of amino acids. Conservative amino acid substitution tables, providing functionally similar amino acids are well known in the art. When referring to nucleic acid

10 sequences, conservative modified variants are those that encode an identical amino acid sequence, in recognition of the fact that codon redundancy allows a large number of different sequences to encode any given protein.

Contig: a series of overlapping cloned sequences e.g. BACs, co-linear and homologous to a region of genomic DNA.

15 **Exons**: protein coding sequences of a gene sequence

Expression cassette: refers to a nucleic acid sequence, comprising a promoter operably linked to a second nucleic acid sequence containing an ORF or gene, which in turn is operably linked to a terminator sequence.

Heterologous: A polynucleotide sequence is "heterologous to" an organism

20 or a second polynucleotide sequence if it originates from a foreign species, or from a different gene, or is modified from its original form. A heterologous promoter operably linked to a coding sequence refers to a promoter from a species, different from that from which the coding sequence was derived, or, from a gene, different from that from which the coding sequence was derived.

25 **Homologue:** is a gene or protein with substantial identity to another gene's sequence or another protein's sequence.

Identity: refers to two nucleic acid or polypeptide sequences that are the same or have a specified percentage of nucleic acids of amino acids that are the same, when compared and aligned for maximum correspondence over a

30 comparison window, as measured using one of the sequence comparison algorithms listed herein, or by manual alignment and visual inspection. When

percentage of sequence identity is used in reference to proteins, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for amino acid residues with similar chemical properties (e.g.,

5   charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to account for the conservative nature of the substitution. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch,

10   thus increasing the percent identity. Means for making these adjustments are well known to those skilled in the art.

Introns: are non-coding sequences interrupting protein coding sequences within a gene sequence.

LCO: lipochitin oligosaccharides.

15   **Legumes:** are members of the plant Family Fabaceae, and include bean, pea, soybean, clover, vetch, alfalfa, peanut, pigion pea, chickpea, fababean, cowpea, lentil in total approximately 20.000 species.

Locus: or "loci" refers to the map position of a nucleic acid sequence or gene on a genome.

20   **Marker assisted breeding**: the use of DNA polymorphisms as "molecular markers", (for examples simple sequence repeats (microsatelittes) or single nucleotide polymorphism (SNP)) which are found at loci, genetically linked to, or within, the *NFR1* or *NFR5* loci, to breed for advantageous *NFR* alleles.

Molecular markers: refer to sites of variation at the DNA sequence level in a

25   genome, which commonly do not show themselves in the phenotype, and may be a single nucleotide difference in a gene, or a piece of repetitive DNA.

Monocotyledenous cereal: includes, but is not limited to, barley, maize, oats, rice, rye, sorghum, and wheat.

Mutant: a plant or organism with a modified genome sequence resulting in a

30   phenotype which differs from the common wild-type phenotype.

Native: as in "native promoter" refers to a promoter operably linked to its homologous coding sequence.

NFR : refers to *NFR* genes or cDNAs, in particular *NFR1* and *NFR5* genes or cDNAs which encode NFR1 and NFR5 polypeptides respectively.

5   **NFR polypeptides**: are polypeptides that are required for Nod-factor binding and function as the Nod-factor binding element in nodulating plants. NFR polypeptides include the NFR5 polypeptide, having an amino acid sequence substantially identical to any one of SEQ ID No: 8, 15, 32, 40 or 48 and the NFR1 polypeptide having an amino acid sequence substantially identical to

10   any one of SEQ ID No: 24, 25, 52 or 54. NFR5 and NFR1 polypeptides show little sequence homology, but they share a similar domain structure comprising an N-terminal signal peptide, an extracellular domain having 2 or 3 LysM-type motifs, followed by a transmembrane domain, followed by an intracellular domain comprising a kinase domain characteristic of

15   serine/threonine kinases. The extracellular domain of NFR proteins is the primary determinant of the specificity of Nod-factor recognition, whereby a host plant comprising a given *NFG* allele will only form nodules with one or a limited number of *Rhizobium* strains. A functional fragment of an NFR polypeptide is one which retains all of the functional properties of a native

20   NFR nod-factor binding polypeptide, including nod-factor binding and interaction with the nod-factor signalling pathway.

Northern blot analysis: a technique for the quantitative analysis of mRNA species in an RNA preparation.

Nod-factors: are synthesised by nitrogen-fixing *Rhizobium* bacteria, which

25   form symbiotic relationships with specific host plants. They are lipo-chitin-oligosaccharides (LCOs), commonly comprising four or five β-1-4 linked N-acetylglucosamines, with a 16 to 18 carbon chain fatty acid n-acetylated on the terminal non-reducing sugar. Nod-factors are synthesised in a number of chemically modified forms, which are distinguished by the compatible host

30   plant.

Nod-factor binding element: comprises one or more NFR polypeptides present in the roots of nodulating plants, and functions in detecting the presence of Nod-factors at the root surface and within the root and nodule tissues. The NFR polypeptides, which are essential for Nod-factor detection,

5   comprise the first step in the Nod-factor signalling pathway that triggers the development of an infection thread and root nodules.

Nod-factor binding properties: are a characteristic of NFR1 and NFR5 polypeptides and are particularly associated with the extracellular domain of said NFR polypeptides, which comprise LysM domains. The binding of Nod-

10   factors by the extracellular domain of NFR polypeptides is specific, such the NFR polypeptides can distinguish between the strain-specific chemically modified forms of Nod-factor.

Nodulating plant: a plant capable of establishing an endosymbiotic *Rhizobium* – plant interaction with a nitrogen-fixing *Rhizobium* bacterium,

15   including the formation of an infection thread, and the development of root nodules capable of fixing nitrogen. Nodulating plants are limited to a few plant families, and are particularly found in the *Legume* family, and they are all member of the Rosid 1 clade.

Non-nodulating plant: a plant which is incapable of establishing an

20   endosymbiotic *Rhizobial* – plant interaction with a nitrogen-fixing *Rhizobial* bacterium, and which does not form root nodules capable of fixing nitrogen.

Operably linked: refers to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

25   **ORF**: Open Reading Frame, which defines one of three putative protein coding sequences in a DNA polynucleotide.

Orthologue: Two homologous genes (or proteins) diverging concurrently with the organism harbouring them diverged. Orthologues commonly serve the same function within the organisms and are most often present in a

30   similar position on the genome.

PCR: Polymerase Chain Reaction is a technique for the amplification of DNA polynucleotides, employing a heat stable DNA polymerase and short oligonucleotide primers, which hybridise to the DNA polynucleotide template in a sequence specific manner and provide the primer for 5' to 3' DNA

5   synthesis. Sequential heating and cooling cycles allow denaturation of the double-stranded DNA template and sequence-specific annealing of the primers, prior to each round of DNA synthesis. PCR is used to amplify DNA polynucleotides employing the following standard protocol or modifications thereof:

10   PCR amplification is performed in 25 µl reactions containing: 10 mM Tris-HCl, pH 8.3 at 25°C; 50 mM KCl; 1.5 mM $MgCl_2$; 0.01% gelatin; 0.5 unit Taq polymerase and 2.5 pmol of each primer together with template genomic DNA (50-100 ng) or cDNA. PCR cycling conditions comprise heating to 94°C for 45 seconds, followed by 35 cycles of 94°C for 20 seconds; annealing at

15   X°C for 20 seconds (where X is a temperature between 40 and 70°C defined by the primer annealing temperature); 72°C for 30 seconds to several minutes (depending on the expected length of the amplification product). The last cycle is followed by heating to 72°C for 2-3 minutes, and terminated by incubation at 4°C.

20   **Pfam consensus**: a consensus sequence derived from a large collection of protein multiple sequence alignments and profile hidden Markov models used to identify conserved protein domains (Bateman *et al.*, 2002, Nucleic Acids Res. 30: 276-80; and searchable on http://www.sanger.ac.uk/Software/Pfam/ and on NCBI at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

25   Protein domain prediction: sequences are analysed by BLAST (www.ncbi.nlm.nih.gov/BLAST/) and PredictProtein (www.embl-heidelberg.de/predictprotein/predictprotein.html). Signal peptides are predicted by SignalP v. 1.1 (www.cbs.dtu.dk/services/signalP/) and transmembrane regions are predicted by TMHMM v. 2.0

30   (www.cbs.dtu.dk/services/TMHMM/)

Polymorphism: refers to "DNA polymorphism" due to nucleotide sequence differences between aligned regions of two nucleic acid sequences.

Polynucleotide molecule: or "polynucleotide", or "polynucleotide sequence" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides

5   and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known analogs of natural nucleotides, which have similar binding properties as the reference nucleic acid.

Promoter: is an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant

10   promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, e.g. a TATA box element, and optionally includes distal enhancer or repressor elements, which can be located several 1000bp upstream of the transcription start site. A tissue specific promoter is one which specifically regulates expressed in a

15   particular cell type or tissue e.g. roots. A "constitutive" promoter is one that is active under most environmental and developmental conditions throughout the plant.

RACE/5'RACE/3'RACE: Rapid Amplification of cDNA Ends is a PCR-based technique for the amplification of 5' or 3' regions of selected cDNA

20   sequences which facilitates the generation of full-length cDNAs from mRNA. The technique is performed using the following standard protocol or modifications thereof: mRNA is reverse transcribed with RNase H⁻ Reverse Transcriptase essentially according to the protocol of Matz *et al*, (1999) *Nucleic Acids Research* 27: 1558-60 and amplified by PCR essentially

25   according to the protocol of Kellogg *et al* (1994) *Biotechniques* 16(6): 1134-7.

Real-time PCR: a PCR-based technique for the quantitative analysis of mRNA species in an RNA preparation. The formation of amplified DNA products during PCR cycling is monitored in real-time, using a specific fluorescent DNA binding-dye and measuring fluorescence emission.

30   **Sexual cross**: refers to the pollination of one plant by another, leading to the fusion of gametes and the production of seed.

SMART consensus: represents the consensus sequence of a particular protein domain predicted by the Simple Modular Architecture Research Tool database (Schultz, J. *et al.* (1998)- *PNAS* 26;95(11):5857-64)

Southern hybridisation: Filters carrying nucleic acids (DNA or RNA) are

5   prehybridized for 1-2 hours at 65°C with agitation in a buffer containing 7 % SDS, 0.26 M Na_2HPO_4, 5 % dextrane-suphate, 1 % BSA and 10µg/ml denatured salmon sperm DNA. Then the denatured, radioactively labelled DNA probe is added to the buffer and hybridization is carried out over night at 65°C with agitation. For low stringency, washing is carried out at 65°C with a

10   buffer containing about 2XSSC, 0.1 % SDS for 20 minutes. For medium stringency, washing is continued at 65°C with a buffer containing about 1XSSC, 0.1 % SDS for 2x 20 minutes and for high stringency filters are washed a further 2x 20 minutes at 65°C in a buffer containing about 0.5XSSC, 0.1 % SDS, or more preferably about 0.3XSSC, 0.1 % SDS.

15   Probe labelling by random priming is performed essentially according to Feinberg and Vogelstein (1983) *Anal. Biochem.* 132(1), 6-13 and Feinberg and Vogelstein (1984) Addendum. *Anal. Biochem.* 137(1), 266-267

Substantially identical: refers to two nucleic acid or polypeptide sequences

20   that have at least about 60%, preferably about 65%, more preferably about 70%, further more preferably about 80%, most preferably about 90 or about 95% nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the sequence comparison algorithms given herein, or by manual alignment and

25   visual inspection. This definition also refers to the complement of the test sequence with respect to its substantial identity to a reference sequence. A comparison window refers to any one of the number of contiguous positions in a sequence (being anything from between about 20 to about 600, most commonly about 100 to about 150) which may be compared to a reference

30   sequence of the same number of contiguous positions after the two sequences are optimally aligned. Optimal alignment can be achieved using

computerized implementations of alignment algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis. USA) or BLAST analyses available on the site: (www.ncbi.nlm.nih.gov/)

5 **TAC clones**: clones from a Transformation-competent Artificial Chromosome library.

TM marker: is a microsatellite marker developed from a TAC sequence, based on sequence differences between *Lotus japonicus* Gifu and MG-20 genotypes.

10 **Transgene**: refers to a polynucleotide sequence, for example a "transgenic expression cassette", which is integrated into the genome of a plant by means other that a sexual cross, commonly referred to as transformation, to give a transgenic plant.

UTR: untranslated region of an mRNA or cDNA sequence.

15 **Variant**: refers to "variant NFR1 or NFR5 polypeptides" encoded by different *NFR* alleles.

Wild type: a plant gene, genotype, or phenotype predominating in the wild population or in the germplasm used as standard laboratory stock.

20 **II. Nod-factor binding**

The present invention provides a Nod-factor binding element comprising one or more isolated NFR polypeptides. The isolated NFR polypeptides, NFR1, as exemplified by SEQ ID No: 24 and 25; and NFR5 (including SYM10) as exemplified by SEQ ID No: 8 and 15 bind to Nod-factors in a chemically-

25 specific manner, distinguishing between the different chemically modified forms of Nod-factors produced by different *Rhizobium* strains. The chemical specificity of Nod-factor binding by NFR1 and NFR5 polypeptides is located in their extracellular domain, which comprises LysM type motifs. The LysM protein motif, first identified in bacterial lysin and muramidase enzymes

30 degrading cell wall peptidoglycans, is widespread among prokaryotes and eukaryotes (Pontig et al. 1999, *J Mol Biol.*289, 729-745; Bateman and

Bycroft, 2000, *J Mol Biol*, 299, 1113-1119). In bacteria it is often found in proteins associated with bacterial cell walls or involved in pathogenesis and *in vivo* and *in vitro* studies of *Lactococcus lactis* autolysin demonstrate that the three LysM domains of this protein bind peptidoglycan (Steen et al, 2003,

5   *J Biol. Chem.* April issue). Since both A- and B-type peptidoglycans, differing in amino acid composition as well as cross-linking were bound, it was concluded that autolysin LysM domains binds the N-acetyl-glucosamine-N-acetyl-murein backbone polymer. LysM domains are frequently found together with amidase, protease or chitinase motifs and two confirmed

10   chitinases carry LysM domains. One is the sex pheromone and wound-induced polypeptide from the alga *Volvox carteri* that binds and degrades chitin *in vitro* (Amon et al.1998,Plant Cell 10,781-9).The other is α-toxin from *Kluyveromyces lactis*, that docs onto a yeast cell wall chitin receptor (Butler,.et al.(1991) Eur J Biochem 199, 483-8). Structure-based alignment

15   of representative LysM domain sequences have shown a pronounced variability among their primary sequence, except the amino acids directly involved in maintaining the secondary structure.

The NFR polypeptides are transmembrane proteins, able to transduce signals perceived by the extracellular NFR domain across the membrane to

20   the intracellular NFR domain comprising kinase motifs, which serves to couple signal perception to the common symbiotic signalling pathway leading to nodule development and nitrogen fixation.

The methods employed for the practise and understanding of the invention, which are described below, involve standard recombinant DNA technology

25   that are well-known and commonly employed in the art and available from Sambrook *et al.*, 1989, *Molecular Cloning: A laboratory manual.*

III. Isolation of nucleic acid molecules comprising *NRF* genes and cDNAs encoding NFR1 and NFR5 polypeptides and their orthologues.

30   The isolation of genes and cDNAs encoding NFR1 or NFR5 (or SYM10) polypeptides, comprising an amino acid sequence substantially similar to

SEQ ID No: 24 or 25 (NFR1); or SEQ ID No: 8 or 15 (NFR5) respectively, may be accomplished by a number of techniques. For instance, a BLAST search of a genomic or cDNA sequence bank of a desired legume plant species (e.g. soybean, pea or *Medicago truncatula*) can identify test

5 sequences similar to the *NFR1* or *NFR5* reference sequence, based on the smallest sum probability score (P(N)). The (P(N)) score (the probability of the match between the test and reference sequence occurring by chance) for a "similar sequence" will be less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001. This approach is

10 exemplified by the *Medicago truncatula* sequence (Ac126779; SEQ ID No: 32) included in Figure 3. Oligonucleotide primers, together with PCR, can be used to amplify regions of the test sequence from genomic or cDNA of the selected plant species, and a test sequence which is similar to the full-length *NFR1* or *NFR5* (or *SYM10*) gene sequences can be assembled. In the case

15 that an appropriate gene bank is not available for the selected plant species, oligonucleotide primers, based on *NFR1* or *NFR5* (or *SYM10*) gene sequences, can be used to PCR amplify similar sequences from genomic or cDNA prepared from the selected plant. The application of this approach is demonstrated in Example 1 A.6, where the isolated *NFR5* gene homologues

20 from *Glycine max* and *Phaseolus vulgaris* are disclosed.

Alternatively, nucleic acid probes based on *NFR1* or *NFR5* (or *SYM10*) gene sequences can be hybridised to genomic or cDNA libraries prepared from the selected plant species using standard conditions, in order to identify clones

25 comprising sequences similar to *NFR1* or *NFR5* genes. A nucleic acid sequence in a library, which hybridises to a *NFR1* or *NFR5* gene-specific probe under conditions which include at least one wash in 2xSSC at a temperature of at least about 65°C for 20 minutes, is potentially a similar sequence to a *NFR1* or *NFR5* (or *SYM10*)gene. The application of this

30 approach is demonstrated in Example 1 B. 4, where the isolation of a pea NFR1 homologue from *Pisum sativum* is disclosed. A test sequence

comprising a full-length cDNA sequence similar to *NFR1* cDNAs having SEQ ID No: 21, or 22, or 51, or 53; or similar to *NFR5* cDNAs having SEQ ID No: 6 or 12 can be generated by 5' RACE cDNA synthesis, as described herein.

5   The nucleic acid sequence of each test sequence, derived from a selected plant species, is determined in order to identify a nucleic acid molecule that is substantially identical to a *NFR1* or *NFR5* gene having SEQ ID No: 23 (*NFR1*), or any one of SEQ ID No: 7, 11, 13, 14, 39 or 47 (*NFR5*) respectively; or a nucleic acid molecule that is substantially identical to a

10   *NFR1* or *NFR5* cDNA having any one of SEQ ID No: 21, 22, 51, or 53 (*NFR1*), or having SEQ ID No: 6 (*NFR5*) or 12 (*SYM10*) respectively; or a nucleic acid molecule that encodes a protein whose amino acid sequence is substantially identical to NFR1 or NFR5 having any one of SEQ ID No: 24, 25, 52 or 54 (NFR1) or having any one of SEQ ID No. 8, 32, 40, or 48

15   (NFR5) or 15 (SYM10), respectively.

IV. Transgenic plants expressing NFR1 and/or NFR5 polypeptides

The polynucleotide molecules of the invention can be used to express a Nod-factor binding element in non-nodulating plants and thereby confer the ability

20   to bind Nod-factors and establish a *Rhizobium*/plant interaction leading to nodule development. An expression cassette comprising a nucleic acid sequence encoding a NFR polypeptide, substantially identical to any one of SEQ ID No: 8, 15, 24, or 25, and operably linked to its own promoter or a heterologous promoter and 3' terminator can be transformed into a selected

25   host plant using a number of known methods for plant transformation. By way of example, the expression cassette can be cloned between the T-DNA borders of a binary vector, and transferred into an *Agrobacterium tumerfaciens* host, and used to infect and transform a host plant. The expression cassette is commonly integrated into the host plant in parallel with

30   a selectable marker gene giving resistance to an herbicide or antibiotic, in order to select transformed plant tissue. Stable integration of the expression

cassette into the host plant genome is mediated by the virulence functions of the *Agrobacterium* host. Binary vectors and *Agrobacterium tumefaciens*-based methods for the stable integration of expression cassettes into all major cereal plants are known, as described for example for rice (Hiei *et al.,*

5   1994, *The Plant J.* 6: 271-282) and maize (Yuji *et al.,* 1996, *Nature Biotechnology,* 14: 745-750). Alternative transformation methods, based on direct transfer can also be employed to stably integrate expression cassettes into the genome of a host plant, as described by Miki *et al.,* 1993, "Procedure for introducing foreign DNA into plants", In: *Methods in Plant Molecular*

10   *Biology and Biotechnology,* Glick and Thompson, eds., CRC Press, Inc., Boca Raton, pp 67-88). Promoters to be used in the expression cassette of the invention include constitutive promoters, as for example the 35S CaMV promoter (Acc V00141 and J02048) or in the case or a cereal host plant the Ubi1 gene promoter (Christensen *et al.,* 1992, *Plant Mol Biol* 18: 675-689). In

15   a preferred embodiment, a root specific promoter is used in the expression cassette, for example the maize *zmGRP3* promoter (Goodemeir et al. 1998, *Plant Mol Biol,* 36, 799.802) or the epidermis expressed maize promoter described by Ponce et al. 2000, *Planta,* 211, 23-33. Terminators that may be used in the expression construct can for instance be the NOS terminator (Acc

20   NC_003065).

Host plants transformed with an expression cassette encoding one NFR polypeptide, for example NFR1, or its orthologue, can be crossed with a second host plant transformed with an expression cassette encoding a second NFR polypeptide, for example NFR5, or its orthologue. Progeny

25   expressing both said NFR polypeptides can then be selected and used in the invention. Alternatively, host plants can be transformed with a vector comprising two expression cassettes encoding both said NFR polypeptides.

V. *NFR* genes encoding NFR polypeptide having specific Nod-factor
30   **binding properties.**

Nucleic acid molecules comprising *NFR1* or *NFR5* genes encoding NFR polypeptides having specific Nod-factor binding properties can be identified by a number of functional assays described in the "Examples" given herein. In a preferred embodiment, said nucleic acid sequences are expressed

5   transgenically in a host plant employing the expression cassettes described above. Expression of *NFR1* or *NFR5* genes or their homologues/orthologues in plant roots allows the specific Nod-factor binding properties of the expressed NFR protein to be fully tested. Assays suitable for establishing specific Nod-factor binding include the detection of: a morphological root hair

10   response (e.g. root hair deformation, root hair curling); a physiological response (e.g. root hair membrane depolarisation, ion fluxes, pH changes and calcium oscillations); a symbiotic signalling response (e.g. downstream activation of symbiotic nodulin gene expression) following root infection with *Rhizobium* bacteria or isolated Nod-factors; the ability to develop root nodule

15   primordia, infection pockets or root nodules, where the response is strain dependent or dependent on the chemical modification of Nod-factor structure.

VI. Marker assisted breeding for *NFR* alleles.

20   A method for marker assisted breeding of *NFR* alleles, encoding variant NFR polypeptides, is described herein, with examples from *Lotus* and *Phaseolus NFR* alleles. In summary, variant NFR1 or NFR5 polypeptides, comprising an amino acid sequence substantially similar to any one of SEQ ID No: 24, 25, 52 or 54 (NFR1) or any one of SEQ ID No: 8, 15, 32, 40 or 48 (NFR5)

25   respectively, are identified in a nodulating legume species, and the *Rhizobium* strain specificity of said variant NRF1 or NFR5 polypeptide is determined, according to measurable morphological or physiological parameters described herein. Subsequently, DNA polymorphisms at loci genetically linked to, or within, the gene locus encoding said variant NFR1 or

30   NFR5 polypeptide, are identified on the basis of the nucleic acid sequence of the loci or its neighbouring DNA region. Molecular markers based on said

DNA polymorphisms, are used for the identification and selection of plants carrying *NFR* alleles encoding said variant NFR1 or NFR5 polypeptides. Use of this method provides a powerful tool for the breeding of legumes with enhanced nodulation frequency.

5

III. Examples

Example 1.

Cloning of Nod-factor Binding Element Genes

Genetic studies in the legume plants *Lotus japonicus* (*Lj*) and pea (*Ps*) have

10   generated collections of symbiotic mutants, which have been screened for mutants blocked in the early steps of symbiosis (Geurts and Bisseling, 2002 *supra*; Kistner and Parniske 2002 *Trends in Plant Science* 7: 511-518). Characteristic for a group of the selected mutants is their inability to respond to Nod-factors, with the absence of root hair deformation and curling, cortical

15   cell division to form the cortical primordium, and induction of the early nodulin genes which contribute to nodule development and function. Nod-factor induced calcium oscillations were also found to be absent in some of these mutants, indicating that they are blocked in an early step in Nod-factor signalling. Among this latter group, are a few mutants, including members of

20   the *Pssym10* complementation group and *LjNFR1* and *LjNFR5* (previously called *Ljsym1 and 5*), which failed to respond to Nod-factors but retain their ability to establish mycorrhizal associations. Genetic mapping indicates that pea *SYM10* and *Lotus NFR5* loci in the pea and *Lotus* could be orthologs. Mutants falling within this group provided a useful starting point in the search

25   for genes encoding potential candidate proteins involved in Nod-factor binding and perception.

A. Isolation, cloning and characterisation of *NFR5* genes and gene products.

1. Map based cloning of *Lj NFR5*

The symbiotic mutants of *Lotus japonicus nfr5-1*, *nfr5-2* and *nfr5-3* (also
known as *sym5*), (previously isolated by Schauser *et al* 1998 *Mo. Gen Genet*,
259: 414-423; Szczglowski *et al* 1998, *Mol Plant-Microbe Interact*, 11: 684-
697) were utilised. To determine the root nodulation phenotype under
symbiotic conditions, seeds were surface sterilised in 2% hyperchlorite,
washed and inoculated with a two day old culture of *M. loti* NZP2235. Plants
were cultivated in the nitrogen-free B&D nutrients and scored after 6-7 weeks
(Broughton and Dilworth, Biochem J, 1971, 125, 1075-1080; Handberg and
Stougaard, *Plant J.* 1992, 2,487-496). Under non-symbiotic conditions, plants
were cultivated in Hornum nutrients (Handberg and Stougaard, *Plant J.* 1992,
2,487-496).

Mapping populations were established in order to localise the *nfr5* locus on
the *Lotus japonicus* genome. Both intra- and interspecific F2 mapping
populations were created by crossing a *Lotus japonicus* "Gifu" *nfr5-1* mutant
to wild type *Lotus japonicus* ecotype "MG20" and to wild type *Lotus
filicaulis*. MG-20 seeds are obtainable from Sachiko ISOBE, National
Agricultural Research Center for Hokkaido Region, Hitsujigaoka, Toyohira,
Sapporo Hokkaido 062-8555, JAPAN and *L. filicaulis* from Jens Stougaard,
Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10,
DK-8000 Aarhus C. F2 plants homozygous for the *nfr5-1* mutant allele were
identified after screening for the non-nodulation mutant phenotype. 240
homozygous F2 mutant plants were analysed in the *L. filicaulis* mapping
population and 368 homozygous F2 mutant plants in the "MG20" mapping
population.

Positional cloning of the *nfr5* locus was performed by AFLP and Bulked
Segregant Analysis of the mapping populations using the *EcoRI/Mse*I
restriction enzyme combination (Vos et al, 1995, *Nucleic Acids Res.*23, 4407-
4414; Sandal et al 2002, *Genetics*, 161, 1673-1683). Initially, *nfr5* was

mapped to the lower arm of chromosome 2 between AFLP markers E33M40-22F and E32M54-12F in the *L. filicaulis* based mapping population, as shown in Figure 1a . The E32M54-12F marker was cloned and used to isolate BAC clones BAC8H12 and BAC67I22 and TAC clone LjT18J10, as

5   shown in Figure 1b. The ends of this contig were used to isolate adjacent BAC and TAC clones namely BAC58K7 and LjT01C03 at one end and TAC LjB06D23 on the other end. The outer end of LjB06D23 was used to isolate TAC clone LjT13I23. The outer end of LjB06D23 was used to isolate TAC clone LjT13I23 (TM0522). Various markers from this contig were mapped on

10   the mapping populations from *nfr5-1* crossed to *L. filicaulis* and to *L. japonicus* MG-20. In the *L. filicaulis* mapping population one recombinant plant was found with the outer end of the TAC clone TM0522, whereas no recombinant plants were found with a marker from the middle of this TAC clone. In the *L. japonicus* MG-20 mapping population, 4 recombinant plants

15   out of 368 plants were found with the marker TM0323, thereby delimiting *nfr5* to a region of 150 kb. This region was sequenced and found to contain 13 ORFs, of which two encoded putative proteins sharing sequence homology to receptor kinases. Sequencing of these two specific ORFs in genomic DNA derived from *nfr5-1* showed that one of the ORF sequences contained a 27

20   nucleotide deletion. Furthermore sequencing of this ORF in genomic DNA from *nfr5-2* and *nfr5-3* showed the insertion of a retrotransposon and a point mutation leading to a premature stop codon, respectively, as shown in Figure 1d. The localisation of the *nfr5* locus from physical and genetic mapping data, combined with the identification of mutations in three independent *nfr5*

25   mutant alleles, provides unequivocal evidence that mutations in the *NFR5* ORF lead to a loss of Nod-factor perception.

2. Cloning the *Lj NFR5* cDNA

A full-length cDNA corresponding to the *NFR5* gene was isolated using a combination of 5´and 3´ RACE. RNA was extracted from *Lotus japonicus*

30   roots, grown in the absence of nitrate or rhizobia, and reverse transcribed to

make a full-length cDNA pool for the performance of 5'-RACE according to the standard protocol. The cDNA was amplified using the 5' oligonucleotide 5'CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT 3' (SEQ ID No:1) and the reverse primer

5 5'GCTAGTTAAAAATGTAATAGTAACCACGC3' (SEQ ID No: 2), and a RACE-product of approximately 2 kb was cloned into a topoisomerase activated plasmid vector (Shuman, 1994, *J Biol Chem* 269: 32678-32684). 3'-RACE was performed on the same 5'-RACE cDNA pool, using a 5' gene-specific primer 5' AAAGCAGCATTCATCTTCTGG 3' (SEQ ID No: 3) and an

10 oligo-dT primer 5'GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTV 3' (SEQ ID No: 4), where the first 5 PCR cycles were carried out at an annealing temperature of 42° C and the following 30 cycles at higher annealing temperature of 58°C. The products of this PCR reaction were used as template for a second PCR reaction with a gene-specific primer positioned

15 further 3' having the sequence 5' GCAAGGGAAGGTAATTCAG 3' (SEQ ID No: 5) and the above oligo dT-primer, using standard PCR amplification conditions (annealing at 54° C; extension 72° C for 30 s) and the products cloned into a topoisomerase activated plasmid vector (Shuman, 1994, *supra*). Nucleotide sequencing of 18 5'RACE clones and three 3' RACE

20 clones allowed the full-length sequence of the *NFR5* cDNA to be determined (SEQ ID No: 6). The NFR5 cDNA was 2283 nucleotides in length, with an open reading frame of 1785 nucleotides, preceded by a 5' UTR leader sequence of 140 nucleotides and a 3'UTR region of 358 nucleotides. Alignment of the *NFR5* cDNA sequence with the *NFR5* gene sequence

25 (SEQ ID No: 7), shown schematically in Figure 1d, confirmed that the gene is devoid of introns.

3. Primary sequence and structural domains of LjNFR5 and mutant alleles.

30 The primary sequence and domain structure of NFR5, encoded by *NFR5,* are consistent with a transmembrane Nod-factor binding protein, required for

Nod-factor perception in rhizobial-legume symbiosis. The *NFR5* gene encodes an NFR5 protein of 596 amino acids having the sequence given in Figure 2b (SEQ ID No: 8) and a predicted molecular mass of 65.3 kD. The protein domain structure predicted for NFR5 and shown in Figure 2a,b,

5   defines a signal peptide, comprising a hydrophobic stretch of 26 amino acids, followed by an extracellular domain with three LysM-type motifs, a transmembrane domain and an intracellular kinase domain. The LysM-type motifs found in *Lotus* NFR5, SYM10, *Medicago truncatula* (*M.t*, Ac126779), and by homology in a rice gene (Ac103891), show homology to the single

10   LysM motif present in an algal (*Volvox carteria*) chitinase (Amon *et al*, 1998, *Plant Cell* 10: 781-789) and to the Pfam consensus, as illustrated in the amino acid sequence alignment of this domain given in Figure 2c. The NFR5 kinase domain has motifs characteristic of functional serine/threonine kinases (Schenk and Snaar-Jagalska, 1999, *Biochim Biophys Acta* 1449: 1-24; Huse

15   and Kuriyan, 2002, *Cell* 109: 275-282), with the exception that motif VII lacks an aspartic acid residue conserved in kinases, and motif VIII, comprising the activation loop, is either divergent or absent.

Analysis of the *nfr5* mutant genes reveals that the point mutation in *nfr5-3* and the retrotransposon insertion in *nfr5-2* will express truncated

20   polypeptides of 54 amino acids, lacking the LysM motifs and entire kinase domain; or of 233 amino acids, lacking the kinase motifs X and XI, respectively. The 27 nucleotide deletion in the *nfr5-1* mutant removes 9 amino acids from kinase motif V.

25   **4. Cloning and characterisation of the pea *SYM10* gene and cDNA and *sym10* mutants.**

Wild type pea cv's (Alaska, Finale, Frisson, Sparkle) and the symbiotic mutants (N15; P5; P56) were obtained from the pea germ-plasm collection at JIC Norwich-UK, while the symbiotic mutant, RisFixG, was obtained from

30   Kjeld Engvild, Risø National Laboratory, 8000 Roskilde, Denmark . The

mutants, belonging to the pea *sym10* complementation group, were identified in the following genetic backgrounds: N15 type strain in a Sparkle background (Kneen *et al,* 1994, *J Heredity 85: 129-133*), P5 in a Frisson background (Duc and Messager, 1989, *Plant Science* 60: 207-213), RisFixG

5   in a Finale background RisFixG (Engvild,1987, *Theoretical Applied Genetics* 74: 711-713; Borisov *et al.*, 2000, *Czech Journal Genetics and Plant Breeding* 36: 106-110); P56 in a Frisson background (Sagan et al.1994, *Plant Science* 100: 59-70).

10   A fragment of the pea *SYM10* gene was cloned by PCR amplification of cv Finale genomic DNA using a standard PCR cycling program and the forward primer 5'-ATGTCTGCCTTCTTTCTTCCTTC-3', (SEQ ID No: 9) and the reverse primer 5'-CCACACATAAGTAATMAGATACT-3', (SEQ ID No: 10). The sequence of these oligonucleotide primers was based on nucleotide

15   sequence stretches conserved in *L. japonicus NFR5* and the partial sequence of an *NFR5* homologue identified in a *M. truncatula* root EST collection (BE204912). The identity of the amplified 551 base pair *SYM10* product was confirmed by sequencing, and then used as a probe to isolate and sequence a pea cv Alaska *SYM10* genomic clone (SEQ ID No:11) from

20   a cv. Alaska genomic library (obtained from H. Franssen, Department of Molecular Biology, Agricultural University, 6703 HA Wageningen, The Netherlands) and a full-length pea cv. Finale *SYM10* cDNA clone (SEQ ID No: 12) from a cv. Finale cDNA library (obtained from H. Franssen, *supra*), which were then sequenced. The sequence of the *SYM10* gene in cv.

25   Frisson (SEQ ID No:13) and in cv. Sparkle (SEQ ID No: 14) were determined by a PCR amplification and sequencing of the amplified gene fragment. The nucleotide sequence of the corresponding mutants P5, P56, and RisFixG were also determined by a PCR amplification and sequencing of the amplified gene fragment.

Nucleotide sequence comparison of the *SYM10* gene in the *Pssym10* mutant lines (P5, RisFix6 and P56) with the wild type parent lines revealed, in each case, sequence mutations, which could be correlated with the mutant phenotype. The 3 independent *sym10* mutant lines identified 3 mutant

5   alleles of the *SYM10* gene, all carrying nonsense mutations, and the N15 type strain was deleted for SYM10 (Table 5). Southern hybridization with probes covering either the extracellular domain of *SYM10* or the 3'UTR on *Eco*RI digested DNA from N15 and the parent variety Sparkle, shows that the *SYM10* gene is absent from the N15 mutant line.

10

5. Primary sequence and structural domains of PsSYM10 and mutant alleles.

The PsSYM10 protein of pea, encoded by *PsSYM10,* is a homologue of the NFR5 transmembrane Nod-factor binding protein of *Lotus*, required for Nod-

15   factor perception in rhizobial-legume symbiosis. The pea cv Alaska *SYM10* gene encodes a SYM10 protein (SEQ ID No: 15) of 594 amino acid residues, with a predicted molecular mass of 66 kD, which shares 73% amino acid identity with the NFR5 protein from *Lotus*. In common with the NFR5 protein, the SYM10 protein has an N-terminal signal peptide, an extracellular region

20   with three LysM motifs, followed by a transmembrane domain, and then an intracellular domain comprising kinase motifs (Figure 2 and 3).
The *sym10* genes in the symbiotic pea mutants P5, RisFix6 and P56, each having premature stop codons, encode truncated SYM10 proteins of 199, 387 and 404 amino acids, respectively, which lack part of, or the entire,

25   kinase domain (Table 5).

6. Isolation of *NFR5* gene orthogues encoding NFR5 protein orthogues

A nucleic acid sequence encoding an NFR5 protein orthologue from bean

30   was isolated from *Phaseolus vulgaris* "Negro jamapa" as follows. A nucleic

acid molecule comprising a fragment of the bean *NFR5* orthologous gene was amplified from *Phaseolus vulgaris* gDNA with the PCR primers:

5'- CATTGCAARAGCCAGTAACATAGA-3' (SEQ ID No: 33) and

5'-AACGWGCWRYWAYRGAAGTMACAAYATGAG-3 (SEQ ID No: 34) using

5   standard PCR reaction conditions (see Definitions: PCR) with an annealing temperature of 48°C, and the amplified fragment was cloned and sequenced. A full-length cDNA molecule corresponding to the amplified bean *NFR5* fragment was obtained by employing 5'-RACE using the oligonucleotide primer: 5'-CGACTGGGATATGTATGTCACATATGTTTCACATG-3' (SEQ ID

10   No: 35) and 3'-RACE using the oligonucleotide primer: 5'-GATAGAATTGCTTACTGGCAGG-3' (SEQ ID No: 36) on bean root RNA according to a standard RACE protocol (see Definitions: RACE). The complete sequence was assembled from both the amplified fragment, 5'RACE - and 3'-RACE products. Finally, the PCR primers: 5'-

15   GACGTGTCCACTGTATCCAGG-3' (SEQ ID No: 37) and 5'-GTTTGGACATGCAATAAACAACTC-3'(SEQ ID No: 38) derived from the assembled sequence, were used to amplify the entire bean *NFR5* gene as a single nucleic acid molecule from genomic DNA of *Phaseolus vulgaris* "Negro Jamapa" and shown to have the sequence of SEQ ID No: 39.

20   A nucleic acid sequence encoding an NFR5 protein orthologue from soybean was isolated from *Glycine max* cv Stevens as follows. A nucleic acid molecule comprising a fragment of the soybean *NFR5* orthologous gene was amplified from *Glycine max* cDNA with the PCR primers:

5'- CATTGCAARAGCCAGTAACATAGA-3' (SEQ ID No: 41) and

25   5'-AACGWGCWRYWAYRGAAGTMACAAYATGAG-3 (SEQ ID No: 42) as described above for the bean *NFR5* orthologue. A full-length cDNA molecule corresponding to the amplified soybean *NFR5* fragment was obtained by employing 5'-RACE using the oligonucleotide primer: 5'-CCATCACTGCACGCCAATTCGTGAGATTCTC -3' (SEQ ID No: 43) and 3'-

30   RACE using the oligonucleotide primer: 5'- GATGTCTTTGCATTTGGGG-3' (SEQ ID No: 44) according to standard protocol (see Definitions: RACE). The

complete sequence was assembled from both the amplified fragment, 5'RACE - and 3'-RACE products. Finally, the PCR primers: 5'-CTAATACGACATACCAACAACTGCAG-3' (SEQ ID No: 45) and 5'-CTCGCTTGAATTTGTTTGTACATG -3'(SEQ ID No: 46) derived from the

5   assembled sequence, were used to amplify the entire soybean *NFR5* gene as a single nucleic acid molecule from genomic DNA of *Glycine max* "Stevens" and shown to have the sequence of SEQ ID No: 48.

Bean *NFR5* gene orthologue from *Phaseolus vulgaris* "Negro jamapa"

10   encodes an NFR5 protein orthologue with an amino acid sequence having SEQ ID No: 40. Soybean *NFR5* gene orthologue from *Glycine max* "Stevens" encodes an NFR5 protein orthologue with an amino acid sequence having SEQ ID No: 48. An alignment of the amino acid sequence of NFR5 orthologues encoded by the *NFR5* gene orthologues isolated from *Lotus*

15   *japonicus, Glycine max* and *Phaseolus vulgaris* is shown in Table 1. All three protein share the common features of three LysM domains, a transmembrane domain and an intracellular protein kinase domain, while kinase domain VII is lacking and domain VIII is highly divergent or absent.

20   The pairwise amino acid sequence similarity between the *Lotus* and *Glycine* NFR5 protein orthologues, and between the *Lotus* and *Phaseolus* NFR5 proteins orthologues is about 80% and about 86 % respectively, while pairwise the nucleic acid sequence similarity between *Lotus NFR5* gene and *Glycine NFR5* and the *Lotus* and *Phaseolus NFR5* gene orthologues is about

25   73% and about 70% respectively (Table 2).

7. The NFR5 protein family is unique to nodulating plants
Comparative analysis defines LjNFR5 and PsSYM10 as members of a novel family of transmembrane Nod-factor binding proteins. A BLAST search of

30   plant gene sequences suggests that genes encoding related, but presently uncharacterised, proteins may be present in the legume *Medicago truncatula*

(Ac126779; figure 2 and 3), while more distantly related, predicted proteins may be found in rice (Ac103891) and *Arabidopsis* (At2g33580), with a sequence identity to NFR5 of 61%, 39%, and 28%, respectively. The high level of sequence conservation in *M. truncatula* (Ac126779) makes this

5 protein and the gene encoding the protein substantially identical to NFR5. In common with the NFR5 and SYM10, the kinase domains of these proteins also lack the conserved aspartic acid residue of motif VII, and the activation loop in motif VIII is highly diverged or absent, as shown in Figure 2d, with the exception of the *Arabidopsis* protein. Only distantly related proteins are

10 therefore found outside the legume family. In conclusion, the NFR5 protein family appears to be restricted to nodulating legumes, and its absence from other plant families may be a key limiting factor in the establishment of rhizobial-root interactions in the members of the families.

15 **8. Tissue specific expression of the *LjNFR5* and *PsSYM10* genes**
The expression pattern of the *NFR5* and *SYM10* genes in *Lotus* and pea is consistent with the role of their gene products as transmembrane Nod-factor binding proteins in the perception of rhizobial Nod-factors at the root surface and later during tissue invasion.

20 The expression of the *NFR5* and *SYM10* genes in various isolated organs of *Lotus* and pea plants, was investigated by determining the steady state *NFR5* and *SYM10* mRNA levels using Real-time PCR and/or Northern blot analysis. Total RNA was isolated from root, leaf, flower, pod and nodule tissues of uninoculated or inoculated *Lotus* "Gifu" or pea plants using a high

25 salt extraction buffer followed by purification through a CsCl cushion. For Northern analysis, according to standard protocols, 20 µg total RNA was size-fractionated on 1.2% agarose gel, transferred to a Hybond membrane, hybridised overnight with an *NFR5* or *SYM10* specific probe covering the extracellular domain and washed at high stringency. Hybridization to the

constitutively expressed ubiquitin *UBI* gene was used as control for RNA loading and quality of the RNA.

For the quantitative real-time RT-PCR, total RNA was extracted using the CsCl method and the mRNA was purified by biomagnetic affinity separation

5   (Jakobsen, K.S. *et al* (1990) *Nucleic Acids Research* 18(12): 3669). The RNA preparations were analysed for contaminating DNA by quantitative PCR and when necessary, the RNA was treated with DNaseI. The DNaseI enzyme was then removed by phenol:chloroform extraction and the RNA was precipitated and re-suspended in 20 µl RNase free H_2O. First strand cDNA

10 was prepared using Expand reverse transcriptase and the quantitative real-time PCR was performed on a standard PCR LightCycler instrument. The efficiency-corrected relative transcript concentration was determined and normalized to a calibrator sample, using *Lotus japonicus* ATP synthase gene as a reference (Gerard C.J. *et al*, 2000 *Mol. Diagnosis* 5: 39-45).

15 The level of *NFR5* mRNA, determined by Northern blot analysis and quantitative RT-PCR, was 60 to 120 fold higher in the root tissue of *Lotus* plants in comparison to other plant tissues (leaves, stems, flowers, pods, and nodules), as shown in Figure 4a. Northern hybridisation show highest expression of *NFR5* in *Lotus* root tissue and a barely detectable expression

20 in nodules. Northern blot analysis detected *SYM10* mRNA in the roots of pea, and a higher level in nodules, but no mRNA was detected in leaves, as shown in Figure 4c.

B. Isolation, cloning and characterisation of *NFR1* genes and gene products.

25 **1. Map based cloning of *Lj NFR1***

The *NFR1* gene was isolated using a positional cloning approach. On the genetic map of *Lotus* the *NFR1* locus is located on the short arm of chromosome I, approximately 22 cM from the top, within a 7.6 cM interval, as

shown in Figure 5a. Several TM markers and PCR markers, derived from DNA polymorphism in the genome sequences of the *L. japonicus* mapping parents, were found to be closely linked to *NFR1* locus and were used to narrow down the region. A physical map of the region, comprising a contig of

5   assembled BAC and TAC clones, is shown in figure 5b. Fine mapping in an F2 population, established from a *Lotus japonicus nfr-1* mutant to wild type *L. japonicus* ecotype 'Miyakojima MG-20' cross, and genotyping of 1603 mutant plants, identified two markers (56K22, 56L2-2) delimiting the *NFR1* locus within a region of 250 kb. BAC and TAC libraries, available from Satoshi

10   Tabata, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0812 Japan; another BAC library from Jens Stougaard, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, DK-8000 Aarhus C, were screened using the closest flanking markers (56L2-1,10M24-1, 36D15) as probes, and the *NFR1* locus was localised to 36 kb within the region. The

15   ORFs detected within the region coded for a UFD1-like protein, a hypothetical protein and a candidate NFR1 protein showing homology to receptor kinases, (Figure 5b).
The region in the genomes of *nfr1-1, nfr1-2* mutants, corresponding to the candidate *NFR1* gene was amplified as three fragments by PCR under

20   standard conditions and sequenced. The fragment of 1827 bp amplified using PCR forward primer 5'TGC ATT TGC ATG GAG AAC C3', (SEQ ID No: 16) and reverse primer 5' TTT GCT GTG ACA TTA TCA GC3', (SEQ ID No: 17) contains single nucleotide substitutions leading to translational stop codons in both the mutant alleles *nfr1-1*, with a CAA to TAA substitution, and the

25   *nfr1-2*, with a GAA to TAA substitution. The physical and genetic mapping of the *nfr1* locus, combined with the identification of mutations in two independent *nfr1* mutant alleles, provides unequivocal evidence that the sequenced *NFR1* gene is required for Nod-factor perception and subsequent signal transduction.

30   **2. Cloning the *Lj NFR1* cDNAs**

Two alternatively spliced *Lj NFR1* cDNAs were identified using a combination of cDNA library screening and 5' RACE on root RNA from *Lotus japonicus*. A *Lotus* root cDNA library (Poulsen *et al.*, 2002, *MPMI* 15:376-379) was screened with an *NFR1* gene probe generated by PCR amplification of the

5   nucleotides between 9689 to 10055 of the genomic sequence, using the primer pair: 5' TTGCAGATTGCACAACTAGG3' (SEQ ID No: 18) and 5'ACTTAGAATCTGCAACTTTGC 3' (SEQ ID No: 19). Total RNA extracted from *Lotus* roots, was amplified by 5' RACE, according to the standard protocol, using the gene specific reverse primer

10   5'ACTTAGAATCTGCAACTTTGC 3' (SEQ ID No 20). Based on the sequence of isolated *NFR1* cDNAs and 5′ RACE products, the *NFR1* gene produces two mRNA species, of 2187 (SEQ ID No: 21) and 2193 nucleotides (SEQ ID No: 22), with a 5' leader sequence of 114 nucleotides, and a 3' untranslated region is 207 nucleotides (Figure 5c). Alignment of genomic and

15   cDNA sequences defined 12 exons in *NFR1* and a gene structure spanning 10235 bp (SEQ ID No: 23). The sequenced region includes 4057bp from the stop codon of the previous gene up to the transcription start point of NFR1 + 6009 bp of NFR1 + 187 bp of 3′genomic. Alternative splice donor sites at the 3′of exon IV account for the two alternative *NFR1* mRNA species.

20

3. Primary sequence and structural domains of LjNFR1 and mutant alleles.

The primary sequence and domain structure of NFR1, encoded by *LjNFR1*, are consistent with a transmembrane Nod-factor binding protein, required for

25   Nod-factor perception in *Rhizobium*-legume symbiosis. The alternatively spliced *NFR1* cDNAs encode NFR1 proteins of 621 (SEQ ID No: 24) and 623 amino acids (SEQ ID No: 25), with a predicted molecular mass of 68.09 kd and 68.23 kd, respectively. The protein has an amino-terminal signal peptide, followed by an extracellular domain having two LysM-type motifs, a

30   transmembrane domain, and an intracellular carboxy-terminal domain comprising serine/threonine kinases motifs

In *nfr1-1*, a stop codon in kinase domain VIII encodes truncated polypeptides of 490 and 492 amino acids, and in *nfr1-2* a stop codon between domain IX and XI encodes truncated polypeptides of 526 and 528 amino acids, as indicated in Figure 6a.

5   In Figure 6b the M1 LysM motif of NFR1 is aligned with the LysM motifs from *Arabidopsis thaliana* and the SMART consensus and M2 LysM of NFR1 with the *Volvox carteri* chitinase (Acc No: T08150), the closest related *Arabidopsis thaliana* receptor kinase (Acc No: NP_566689) , the rice (Acc No: BAB89226) and the consensus SMART LysM motif.

10

4. Isolation of *NFR1* gene orthogues encoding NFR1 protein orthogues

Two nucleic acid molecules have been isolated from a *Pisum sativum* cv Finale (pea) root hair cDNA library, that comprise two cDNA molecules encoding NFR1 A and NFR1B protein orthologues. The pea cDNA library

15   was screened by hybrisation at medium stringency (see Definitions: Southern hybridisation) using a *Lotus NFR1* gene probe, comprising the coding region for the extracellular domain of *Lotus* NFR1. This *NFR1* gene specific probe was amplified from the *Lotus NFR1* coding sequence by PCR using the primers: 5´- TAATTATCAGAGTAAGTGTGAC-3´ (SEQ ID No: 49) and 5´-

20   AGTTACCCACCTGTGGTAC-3´ (SEQ ID No. 50).

The two cDNA clones *Pisum sativum NFR1A* (SEQ ID No: 51) and *Pisum sativum NFR1B* (SEQ ID No: 53) encode the orthologues NFR1A (SEQ ID No: 52) and NFR1B (SEQ ID No: 54) respectively. An alignment of the amino

25   acid sequence of the three NFR1 orthologues from *Lotus* and *Pisum sativum* is shown in Table 3. All three protein share the common features of LysM domains, a transmembrane domain and an intracellular protein kinase domain, while kinase domain VII is lacking and domain VIII is highly divergent or absent. The nucleic acid sequence of the *Pisum* and *Lotus*

30   *NFR1* orthologues show close similarity (about 83%), as do their respective encoded proteins (about 73%) as shown in Table 4.

4. The LjNFR1 protein family is not found in non-nodulating plants

Comparative analysis defines LjNFR1 as a member of a second novel family of transmembrane Nod-factor binding proteins. Although proteins having both
5   receptor-like kinase domains and LysM motifs are predicted from plant genome sequences, their homology to NFR1 is low and their putative function unknown. *Arabidopsis* has five predicted receptor-like kinases with LysM motifs in the extracellular domain, and one of them (At3g21630) is 54% identical to NFR1 at the protein level. Rice has 2 genes in the same class,
10   and one (BAB89226) encodes a protein with 32 % identity to NFR1.
This suggests that the NFR1 protein is essential for Nod-factor perception and its absence from non-nodulating plants may be a key limiting factor in the establishment of rhizobial-root interactions in these plants. Although NFR1 shares the same domain structure to NFR5 their primary sequence homology
15   is low (Figure 11).

5. Expression of the *LjNFR1, NFR5 and SymRK* symbiotic genes is root specific and independently regulated.

The *NFR1* dependent root hair curling, in the susceptible zone located just
20   behind the root tip, is correlated with root specific *NFR1* gene expression. Steady-state *NFR1* mRNA levels were measured in different plant organs using quantitative real-time PCR and Northern blot analysis as described above in section A.7. *NFR1* mRNA was only expressed in root tissue, and remained below detectable levels in leaves, flowers, pods and nodules, as
25   shown in Figure 7a. Upon inoculation with *M. loti*, the expression of *NFR1* in wild type plants is relatively stable for at least 12 days after inoculation (Figure 7b). Real-time PCR experiments revealed no difference between the levels of the two *NFR1* transcripts detected in the root RNA, suggesting that the alternative splicing of exon 4 is not differentially regulated.
30   *NFR1, NFR5* and *SymRK* gene expression in roots, before and following *Rhizobium* inoculation, was determined by real-time PCR in wild type and

nfr1, nfr5 and symrk mutant genotypes. The expression of *NFR1, NFR5* and *SymRK* genes in un-inoculated and inoculated roots was not significantly influenced by the symbiotic mutant genotype (Figure 7b, c, d) indicating that transcriptional regulation of these genes is mutually independent.

5

Example 2.

Functional properties of the Nod-factor binding element and its component NFR proteins

The functional and regulatory properties of the Nod-factor binding element
10 and its component NFR proteins provide valuable tools for monitoring the functional expression and specific activity of the NFR proteins. Nod-factor perception by the Nod-factor binding element triggers the rhizobial-host interaction, which includes depolarisation of the plasma membrane, ion fluxes, alkalization of the external root hair space of the invasion zone,
15 calcium oscillations and cytoplasmic alkalization in epidermal cells, root hair morphological changes, infection thread formation and the initiation of the nodule primordia. These physiological events are accompanied and coordinated by the induction of specific plant symbiotic genes, called nodulins. For example, the *NIN* gene encodes a putative transcriptional
20 regulator facilitating infection thread formation and inception of the nodule primordia and limits the region of root cell-rhizobial interaction competence to a narrow invasion zone (Geurts and Bisseling, 2002, *supra*). Since *nin* mutants develop normal mycorrhiza, the *NIN gene* lies in the rhizobia-specific branch of the symbiotic signalling pathway, downstream of the common
25 pathway. Ion fluxes, pH changes, root hair deformation and nodule formation are all absent in *NFR1* and *NFR5* mutant plants, and hence the functional activity of these genes must be required for all downstream physiological responses. Several physiological and molecular markers that are diagnostic of *NFR* expression are provided below.

30

1. Morphological marker of *NFR1* and *NFR5* gene expression

When wild type *Lotus japonicus* plants are inoculated with *Mesorhizobium loti*, the earliest visible evidence of infection is root hair deformation and root hair curling, which occurs 24 hours after inoculation, as shown in Figure 8a.

5   However, mutant plants carrying the *nfr1-1* (Figure 8c), *nfr1-2*, *nfr5-1*, *nfr 5-2* or *nfr5-3* alleles (as in Figure 8c), all failed to produce root hair curling or deformation, infection threads or nodule primordia in response to infection by *Mesorhizobium loti* with all three strains tested (NZP2235, R7A and TONO). Lipochitin-oligosaccharides purified from *M. loti*, R7A strain, which induce

10   root hair deformation and branching in wild type plants (Figure 8b), also failed to induce any deformation of root hairs of the *nfr1-1* and *nfr5-1* mutants (Figure 8d), evidencing the key role of the *NFR1* and *NFR5* genes in Nod-factor perception.

Mutations in genes expressing the downstream components of the symbiosis

15   signalling pathway, namely *symRK and nin* have clearly distinguishable phenotypes. After infection with *Mesorhizobium loti*, the root hairs of *symRK* plants swell into balloon structures (Figure 8e), while the *nin* mutants produce an excessive root hair response (Figure 8g). The response of double mutants carrying *nfr1-1/symRK-3* mutant alleles or *nfr1-1/ nin* alleles to

20   *Mesorhizobium loti* infection (Figure 8f,h) are similar to that of *nfr1-1* mutants, demonstrating that the *nfr1-1* mutation is dominant to *symRK* and *nin* mutations, and hence determines an earlier step in the symbiotic signalling pathway.

25   ### 2. Physiological marker of *NFR1* and *NFR5* gene expression

When the root hairs of wild type *Lotus* plants are exposed to *M. loti* Nod-factor, the plasma membrane is depolarised and an alkalisation occurs in the root hair space of the invasion zone, (Figure 9a). The extracellular pH was monitored continuously in a flow-through regime using a pH-selective

30   microelectrode, placed within the root hair space. Membrane potential was measured simultaneously with pH, and the calculated values are based on at

least three equivalent experiments, each. Mutants carrying *nfr1* and *nfr5* alleles do not respond normally to Nod-factor stimulation. Two *nfr5* alleles abolish the response to Nod-factors (Figure 9b), while the *nfr1-1* allele causes a diminished and slower alkalisation, and the *nfr1-2* allele causes the

5 acidification of the extracellular root hair space (Figure 9c). Both the *NFR1* and *NFR5* genes are thus essential for mounting the earliest detectable cellular and electrophysiological responses to Nod-factor, which can be used to monitor their functional activity.

The early physiological response of the *symRK-3 and symRK-1* mutant

10 plants to *Mesorhizobium loti* Nod-factor is similar to the wild type (Figure 9d) and clearly distinguishable from the response of both the *nfr*1 and *nfr5* mutants.

The response of the double mutant, carrying *nfr1-2/symRK-3* mutant alleles, to Nod-factor (Figure 9e) is similar to that of *nfr1-2* mutants, further

15 supporting that the *nfr1-2* mutation is dominant to *symRK-3* and determines an earlier step in the symbiotic signalling pathway.

3. *NFR1* and *NFR5* mediated Nod-factor perception lies upstream of *NIN* and *ENOD* and is required for their expression.

20 The symbiotic expression of the nodulin genes, *Lotus japonicus ENOD2* (Niwa, S. *et al.*, 2001 *MPMI* 14:848-56) and *NIN*, in roots following rhizobial inoculation, provides a marker for *NFR* gene expression. The steady-state levels of *NIN* and *ENOD2* mRNA were measured in roots before and following rhizobial inoculation by quantitative real-time PCR, using the primer

25 pairs:
5'AATGCTCTTGATCAGGCTG3' (SEQ ID No: 26) and
5'AGGAGCCCAAGTGAGTGCTA3' (SEQ ID No: 27) for amplification of *NIN* mRNA reverse transcripts; and the primer pairs:
5'CAG GAA AAA CCA CCA CCT GT3' (SEQ ID No:28) and

5'ATGGAGGCGAATACACTGGTG3' (SEQ ID No: 29) for amplification of *ENOD2* mRNA reverse transcripts. The identity of the amplified sequences was confirmed by sequencing.

Five hours after inoculation, induction of *NIN* gene expression was detected
5   in the wild type plants, while induction of *ENOD2* occurs after 12 days as shown in Figure 10a and b. In the *nfr1* and *nfr5* mutants, activation of *NIN* and *ENOD2* was not detected, demonstrating that functional *NFR1* and *NFR5* genes can be monitored by the activation of these early nodulin genes. *Lotus* plants transformed with a *NIN* gene promoter region fused to a GUS
10 reporter gene provide a further tool to monitor *NFR* gene function. Expression of the *NIN*-GUS reporter can be induced in root hairs and epidermal cells of the root invasion zone following rhizobial inoculation in transformed wild-type plants. In contrast expression of the *NIN*-GUS reporter in an *nfr1* mutant was not detected following rhizobial inoculation. Likewise,
15 *NIN*-GUS expression was induced in the invasion zone of wildtype plants after Nod-factor application, while in a *nfr1* mutant background no expression was detected The requirement for *NFR1* function was confirmed in *nfr1-1*, *nin* double mutants by the absence of root hair curling and excessive root hair curling (Fig 8).

20 The *LjCBP1* gene, T-DNA tagged with a promoter-less GUS in the T90 line, is rapidly activated after *M. loti* inoculation as seen for *NIN*-GUS, thus providing an independent and sensitive reporter of early nodulin gene expression (Webb et al, 2000, Molecular Plant-Microbe Interact. 13,606,-616). Parallel experiments comparing expression of the *LjCBP1* promoter
25 GUS fusion in wt and *nfr1* mutant background confirm the requirement for a functional *NFR1* for activation of the early response to bacteria and Nod-factor.

Example 3.

Transgenic expression of NFR polypeptides and complementation of the *nfr* mutants

The *NFR* genes, encoding the NFR1 and NFR5 protein components of the Nod-factor binding element, can each be stabily integrated, as a transgene,

5   into the genome of a plant, such as a non-nodulating plant or a mutant non-nodulating plant, by transformation. Expression of this transgene, directed by an operably linked promoter, can be detected by expression of the respective NFR protein in the transformed plant and functional complementation of a non-nodulating mutant plant.

10   A wildtype *NFR5* transgene expression cassette of 3,5 kb, comprising a 1175 bp promotor region, the *NFR5* gene and a 441 bp 3' UTR was cloned in a vector (pIV10), and the vector was recombined into the T-DNA of *Agrobacterium rhizogenes* strain AR12 and AR1193 by triparental mating. The *NFR5* expression cassette in pIV10 was subsequently transformed into

15   non-nodulating *Lotus nfr5-1* and *nfr5-2* mutants via *Agrobacterium rhizogenes*-mediated transformation according to the standard protocol (Stougaard 1995, Methods in Molecular Biology volume 49, Plant Gene Transfer and Expression Protocols, p 49-63) In parallel, control transgenic *Lotus nfr5-1* and *nfr5-2* mutants plants were generated, which were

20   transformed with an empty vector, lacking the *NFR5* expression cassette. The nodulation phenotype of the transgenic hairy root tissue of the transformed mutant *Lotus* plants was scored after inoculation with *Mesorhizobium loti* (*M. loti*) strain NZP2235. *In planta* complementation of the *nfr5-1* and *nfr5-2* mutants by the *NFR5* transgene was accomplished, as

25   shown in Table 6, with an efficiency of ≥58%, and the establishment of normal rhizobial-legume interactions and development of nitrogen fixing nodules. Complementation was dependent on transformation with a vector comprising the *NFR5* expression cassette.

A transgene expression cassette, comprising the wild type *NFR1* gene

30   comprising 3020 bp of promoter region, the *NFR1* ORF and 394 bp of 3'untranslated region, was cloned into the pIV10 vector and recombined into

Agrobacterium rhizogenes strain AR12 and AR1193 by triparental mating. *Agrobacterium rhizogenes*-mediated transformation was used to transform the gene into non-nodulating *Lotus nfr1-1* and *nfr1-2* mutants in parallel with a control empty vector. *In planta* complementation of the *Lotus nfr1-1* and

5   *nfr1-2* mutants by the *NFR1* transgene was accomplished, as shown in Table 7, with an efficiency of ≥60%, and the establishment of normal *Rhizobium*-legume interactions with *M. loti* strain NZP2235, and development of nitrogen fixing nodules. Complementation was dependent on transformation with a vector comprising the *NFR1* expression cassette

10   **Example 4**

Expression and characterisation of the NFR1, NFR5 and SYM10 proteins in transgenic plants

NFR1, NFR5 and SYM10 proteins are expressed and purified from transgenic plants, by exploiting easy and well described transformation

15   procedures for *Lotus* (Stougaard 1995, *supra*) and tobacco (Draper et al.1988, Plant Genetic Transformation and Gene Expression, A Laboratory Manual, Blackwell Scientific Publications). Expression in plants is particularly advantageous, since it facilitates the correct folding of these transmembrane proteins and provides for correct post-translational modification, such as

20   phosphorylation. The primary sequences of the expressed proteins are extended with commercially available epitope tags (Myc or FLAG), to allow their purification from plant protein extracts. DNA sequences encoding the tags are ligated into the expression cassette for each protein, in frame, either at the 5′ or the 3′ end of the cDNA coding region. These modified coding

25   regions are then operably linked to a promoter, and recombined into *Agrobacterium rhizogenes*. *Lotus* is transformed by wound-site infection and from the transgenic roots independent root cultures are established *in vitro* (Stougaard 1995, *supra*). NFR1, NFR5 and SYM10 proteins are then purified

from root cultures by affinity chromatography using the epitope specific antibody and standard procedures. Alternatively the proteins are immunoprecipitated from crude extracts or from semi-purified preparations. Proteins are detected by Western blotting methods. For transformation and

5 expression in tobacco, the epitope tagged cDNAs are cloned into an expression cassette comprising a constitutively expressed 35S promoter and a 3'UTR and subsequently inserted into binary vectors. After transfer of the binary vector into *Agrobacterium tumefaciens*, transgenic tobacco plants are obtained by the transformation regeneration procedure (Draper et al.1988,

10 *supra*). Proteins are then extracted from crude or semi-purified extracts of tobacco leaves using affinity purification or immunoprecipitation methods. The epitope tagged purified protein preparations are used to raise mono-specific antibodies towards the NFR1, NFR5 and SYM10 proteins

15 **Example 5**

Plant breeding tools to select for enhanced nodulation frequency and efficiency.

A successful and efficient primary interaction between a rhizobial strain and its host depends on detection of a *Rhizobium* strain's unique Nod-factor

20 (LCO) profile by the plant host. The Nod-factor binding element and its component NFR proteins, each with their extracellular LysM motifs, play a key role in controlling this interaction. *NFR* alleles, encoding variant NFR proteins are shown to be correlated with the efficiency and frequency of nodulation with a given rhizobial strain. Molecular breeding tools to detect

25 and distinguish different plant *NFR* alleles, and assays to assess the nodulation efficiency and frequency of each allele, provides an effective method to breed for nodulation efficiency and frequency.

Methods useful for breeding for nodulation efficiency and frequency are given below, and the application of these techniques is illustrated for the *NFR* alleles of *Lotus* spp. Using the *Rhizobium leguminosarum* bv *viceae* 5560 DZL strain (Bras et al, 2000, Molecular Plant-Microbe Interact. 13, 475-479) it

5   is documented that the host range of this strain within the *Lotus* spp depends on the *NFR1* and *NFR5* alleles present in the *Lotus* host. When inoculated onto wild type plants *Rhizobium leguminosarum* bv *viceae* 5560 DZL form root nodules on *Lotus japonicus* GIFU but the strain is unable to form root nodules on *Lotus filicaulis*. Transgenic *L. filicaulis* transformed with the *Lotus*

10   *japonicus* GIFU *NFR1* and *NFR5* alleles do however form root nodules when inoculated with the *Rhizobium leguminosarum* bv *viceae* 5560 DZL strain proving the NFR1/NFR5 allele dependent Nod-factor recognition.

1. Determining the Nod-factor specificity and sensitivity of *NFR alleles*.

Root hair curling and root hair deformation in the susceptible invasion zone is

15   a sensitive *in vivo* assay for monitoring the legume plants ability to recognise a *Rhizobium* strain or the Nod-factor synthesized by a *Rhizobium* strain. The assay is performed on seedlings and established as follows. Seeds of wild type, transgenic and mutant *Lotus* spp are sterilised and germinated for 3 days. Seedlings are grown on 1/4 B&D medium (Handberg and Stougaard,

20   1992 *supra*), between two layers of sterile wet filter paper for 3 days more. Afterwards, they are transferred into smaller petri dishes containing 1/4 B&D medium supplemented with 12.7nM AVG [(S)-trans-2-amino-4-(2-aminoethoxy)-3-butenoic acid hydrochloride] (Bras C. et al, 2000 , *MPMI* 13: 475-479). On transfer, the seedlings are inoculated with either 20 µl of 1:100

25   dilution of a 2 days old *M.loti* strain NZP2235 culture, or with *M.loti* strain R7A Nod-factor coated sand, or with sterile water as a control, and a layer of wet dialysis membrane is used to cover the whole root. A minimum of 30 seedlings are microscopically analysed for specific deformations of the root hairs. The assay determines the threshold sensitivity of each *L. japonicus* ,

for the Nod-factor (LCO) of a given *Rhizobium* strain and the frequency of root hair curling and/or deformation.

In an alternative procedure, seeds of *Lotus japonicus* are surface sterilised and germinated for 4 days on 1% agar plates containing half-strength

5   nitrogen-free medium (Imaizumi-Anraku *et al.*, 1997, *Plant Cell Physiol.* 38: 871-881), at 26°C, under a 16h light and 8h dark regime. Straight roots, of <1cm in length, on germlings from each cultivar are then selected and transplanted on Fåhraeus slides, in a nitrogen-free medium and grown for a further 2 days. LCOs, prepared by *n*-butanol extraction and HPLC separation

10   from a given *Rhizobium* strain (Niwa *et al.* 2001, *MPMI* 14: 848-856), are applied to the straight roots in each cultivar, at a final concentration range of between 10^{-7} and 10^{-9} M. After 12 to 24h culture, the roots are stained with 0.1% toluene blue and the number of root hairs showing curling is counted. The assay determines the threshold sensitivity of each *Lotus* spp., carrying a

15   given *NFR* allele, for the Nod-factor (LCO) of a given *Rhizobium* strain and the frequency of root hair curling.

2. Determining the frequency and efficiency of nodulation of *NFR* alleles.

The efficiency of a legume plants ability to form root nodules after inoculation

20   with a *Rhizobium* strain is determined in small scale controlled nodulation tests. *Lotus* seeds are surface sterilised in 2 % hyperchlorite and cultivated under aseptic conditions in nitrogen free 1/4 concentrated B&D medium. After 3 days of germination, seedlings are inoculated with a 2 days old culture of *M. loti* NZP2235 or TONO or R7A or with the *R. leguminosarum* bv

25   *viceae* 5560DZL strain. In principle a set of plants is only inoculated with one stain. For controlled competition experiments where legume-*Rhizobium* recognition is determined in a mixed *Rhizobium* population, a set of plants can be inoculated with more than one *Rhizobium* strain or with an extract

from a particular soil. Two growth regimes are used: either petri dishes with solidified agar or Magenta jars with a solid support of burnt clay and vermiculite. The number of root nodules developed after a chosen time period is then counted, and the weight of the nodules developed can be

5   determined. The efficiency of the root nodules in terms of nitrogen fixation can be determined in several ways, for example as the weight of the plants or directly as the amount of N15 nitrogen incorporated in the plant molecules.

In an alternative procedure, *Lotus* seeds are surface sterilised and vernalised at 4°C for 2 days on agar plates and germinated overnight at 28°C. The

10   seedlings are inoculated with *Mesorhizobium loti* strain NZP2235, TONO or R7A LCOs (as described above) and grown in petri dishes on Jensen agar medium at 20°C in 8h dark, 16h light regime. The number of nodules present on the plant roots of each cultivar is determined at 3 days intervals over a period of 25 days, providing a measure of the rate of nodulation and the

15   abundance of nodules per plant.

3. Determining nodule occupancy in relation to *NFR* allele

In agriculture the NFR Nod-factor binding element recognises *Rhizobium* bacteria under adverse soil conditions. The final measure of a particular strain´s or commercial *Rhizobium* inoculum´s ability to compete with the

20   endogenous *Rhizobium* soil population for invasion of a legume crop with particular NFR alleles, is root nodule occupancy. The proportion of nodules formed after invasion by a particular strain and the fraction of the particular *Rhizobium* strain inside individual root nodules is determined by surface sterilising the root nodule surface in hyperchlorite, followed by crushing of the

25   nodule into a crude extract and counting the colony forming *Rhizobium* units after dilution of the extract and plating on medium allowing *Rhizobium* growth (Vincent., JM. 1970, A manual for the practical study of root nodule bacteria. IBP handbook no. 15 Oxford Blackwell Scientific Publications, López-García

et al, 2001, *J Bacteriol*, 183,7241-7252).

4. *NFR1* and *NFR5* are determinants of host range in *Lotus-Rhizobium* interactions.

5   Wild type *Lotus japonicus* Gifu is nodulated by both *Rhizobium leguminosarum bv. viciae* 5560 DZL (*R. leg* 5560DZL) and *Mesorhizobium loti* NZP2235 (*M.loti* NZP2235), while wild type *Lotus filicaulis* is only nodulated by *M.loti* NZP2235. Transgenic *Lotus filicaulis* plants expressing the *NFR1* and *NFR5* alleles of *Lotus japonicus* Gifu, are nodulated by *R. leg*

10   5560DZL, clearly demonstrating that the *NFR* alleles are the primary determinants of host range.

Lotus filicaulis was transformed with vectors comprising *NFR1* and *NFR5* wild type genes and their cognate promoters from *Lotus japonicus* Gifu or with empty vectors. The *Lotus filicaulis* transformants carrying *NFR1* and

15   *NFR5* are nodulated by *R. leg* 5560DZL, albeit at reduced efficiency/frequency (9.6%) compared to *Lotus japonicus* Gifu (100%), as shown in Table 8. Mixing of NFR subunits from *Lotus japonicus* and *Lotus filicaulis* in the Nod-factor binding element is likely to contribute to the reduced efficiency observed. These data demonstrate that rhizobial strain

20   recognition specificity is determined by the *NFR1* and *NFR5* alleles and that breeding for specific *NFR* alleles present in the germplasm or in wild relatives can be used to select optimal legume-*Rhizobium* partners.

More detailed investigations show that the rhizobial strain recognition

25   specificity of the *NFR5* and *NFR1* alleles is determined by the extracellular domain of the NFR5 and NFR1 proteins. Mutant *Lotus japonicus nfr5* was transformed with a wild type hybrid *NFR5* gene "FinG5", encoding the extracellular domain from *L. filicaulis* NFR5 fused to the kinase domain from

L. japonicus Gifu NFR5 (Figure 12). The hybrid gene was operably linked to the wild type *NFR5* promoter. Control transformants, comprising wild type *L. japonicus* Gifu, L. *filicaulis* and the *Lotus japonicus nfr5* mutant, transformed with an empty vector, are generated in parallel. The transformed plants are

5   infected either with *M.loti* NZP2235 or with *R. leg*5560 DZL and the formation of nodules monitored, as shown in Table 9. The FinG5 hybrid gene complements the *nfr5* mutation, and 88% of the transformants are nodulated by *M.loti* NZP2235 showing that the hybrid gene is functionally expressed. However, the *nfr5* mutants expressing the FinG5 hybrid gene are very poorly

10   nodulated by *R.leg* 5560 DZL, only 3 %, (corresponding to one plant) even after prolonged infection (40 days). This demonstrates that strain specificity of the Nod-factor binding element is determined by the extracellular domain of its component NFR proteins.

15   In parallel, the *Lotus japonicus nfr1* mutant was transformed with a wild type hybrid *NFR1* gene "FinG1", encoding the extracellular domain from *L. filicaulis* NFR1 fused to the kinase domain from *L. japonicus* Gifu NFR1 (Figure 12). The hybrid gene was operably linked to the wild type *NFR1* promoter. The transformed plant were infected either with *M.loti* NZP2235 or

20   with *R. leg* 5560 DZL and the formation of nodules was monitored, as shown in Table 10.
The FinG1 hybrid gene complements the *nfr1-1* mutation, and 100 % of the transformants were nodulated by *M.loti* NZP2235. However *nfr1-1* mutants expressing the FinG1 hybrid gene were less efficiently nodulated (30-40%)

25   by *R. leg* 5560 DZL. Furthermore, their nodulation by *R. leg* 5560 DZL was much delayed compared to their nodulation by *M. loti* NZP2235. Thus the *Lotus / R. leg* 5560 DZL interaction is less efficient and delayed when the transgenic host plant expresses a hybrid NFR1 comprising the extracellular domain of *Lotus filicaulis* NFR1 with the kinase domain of *Lotus japonicus*

30   Gifu NFR1. These data indicate that the specific recognition of *R.leg 5560* DZL by its *Lotus* host is at least partly specified by the extracellular domain of

NFR1 (Gifu) and that this is an allele specific recognition. However, the *NFR5* allele appears to be more important for specific recognition than *NFR1*.

5. *NFR5* and *NFR1* alleles and their molecular markers

5   The NFR5 Nod-factor binding proteins encoded by the *NFR5* alleles of *Lotus japonicus* ecotype GIFU (gene sequence: SEQ ID No: 7; protein sequence: SEQ ID No: 24 & 25), and *Lotus filicaulis* (gene sequence SEQ ID No: 30; protein sequence SEQ ID No: 31) have been compared, and found to show diversity in their primary structure. Using the sequence information available

10   for the *Lotus NFR5* gene together with the pea SYM10 gene (Table 12), the alleles from different ecotypes or varieties of *Lotus,* pea and other legumes can now be identified, and used directly in breeding programs. By further way of example, the nucleic acid sequence of the *Phaseolus vulgaris NFR5* gene (SEQ ID No: 39) has facilitated the identification of a molecular marker for

15   two different *NFR5* alleles in the *Phaseolus vulgaris* lines Bat93 and Jalo EEP558, that is based on a single nucleotide difference creating an *ApoI* restriction site (RAATTY) in line Bat93, wherein R stands for A or G, Y for C or T. A partial sequence of the *NFR5* gene comprising the *ApoI* site molecular marker identified in line Bat93 is shown in bold type:

20   CACAGGACATATTGAGTGAAAACAACTATGGTCA**AAATTT**CACTGCCGC AAGCAACCTTCCAGTTTTGATCCCAGTTACA
The absence of this *ApoI* site in the comparable *NFR5* partial sequence of line Jalo EEP558 is shown in bold type:
CACAGGACATATTGAGTGAAAACAACTATGGTCA**AAACTT**CACTGCCGC

25   AAGCAACCTTCCAGTTTTGATCCCAGTTACA
Molecular markers based on DNA polymorphism are used to detect the alleles in breeding populations. Similar use can be taken of the *NFR1* sequences. Molecular DNA markers, based on the *NFR5* allele sequence differences of *Lotus* and pea, are highlighted in Tables 12 and 13 as

examples of how DNA polymorphism can be used directly to detect the presence of an advantageous allele in a breeding population.

Breeding for an advantageous allele can also be carried out using molecular markers, that are genetically linked to the allele of interest, but located

5 outside the gene-allele itself. Breeding of new *Lotus japonicus* lines containing a desired *NFR5* allele can, for example, be facilitated by the use of DNA polymorphisms, (simple sequence repeats (microsatelittes) or single nucleotide polymorphism (SNP) which are found at loci, genetically linked to *NFR5*. Microsatelittes and SNPs at the *NFR5* locus are identified by

10 transferring markers from the general map, by identification of AFLP markers, or, by scanning the nucleotide sequence of the BAC and TAC clones spanning the *NFR5* locus, for DNA polymorphic sequences located in close proximity of the *NFR5* gene. Table 11 lists the markers closely linked to *NFR5* and the sequence differences used to design the microsatelitte or SNP

15 markers. This principle of marker assisted breeding, using genetically linked markers, can be applied to all plants. Microsatellite markers which generate PCR products with a high degree of polymorphism, are particularly useful for distinguishing closely related individuals, and hence to distinguish different *NFR5 of NFR1* alleles in a breeding program.

Table 1

Alignment of *Lotus*, *Glycine* and *Phaseolus* NFR5 protein sequences

```
                       1          2          3          4          50
Lotus      1  MAVFF--    GSLSLFLALT  LLFTNIAARS  EKISGPDFSC  PVDSPPSCET    50
Glycine    1  MAVFPPFLPL  HSQILCLVIM  LFSTNIVAQS  QQDNRTNFSC  PSDSPPSCET    50
Phaseolus  1  MAVFFVSLTL  GAQILYVVLM  FFTC-       QQTNGTNFSC  PSNSPPSCET    50

                       6          7          8          9          100
Lotus      5  YVTYTAQSPN  LLSLTNISD   FDISPLSIA   ASNIDAGKDK  LVPGQVLLVP   100
Glycine    5  YVTYIAQSPN  FLSLTNISN   FDTSPLSIAR  ASNLEPMDDK  LVKDQVLLVP   100
Phaseolus  5  YVTYISQSPN  FLSLTSVSN   FDTSPLSIAR  ASNLQHEEDK  LIPGQVLLI    100

                       11         12         13         14         150
Lotus      10 VTCGCAGNHS  SANTSYQIQL  GDSYDFVATT  LYENLTNWNI  VQASNPGVNP   150
Glycine    10 VTCGCTGNRS  FANISYEINQ  GDSFYFVATT  SYENLTNWRA  VMDLNPVLSP   150
Phaseolus  10 VTCGCTGNRS  FANISYBINQ  GDSFYFVATT  LYQNLTNWHA  VMDLNPGLSQ   150

                       16         17         18         19         200
Lotus      15 YLLPERVKVV  PPLFCRCPSK  NQLNKGIQYL  ITYVWKPNDN  VSLVSAKFGA   200
Glycine    15 NKLPIGIQVV  FPLFCKCPSK  NQLDKEIKYL  ITYVWKPGDN  VSLVSDKFGA   200
Phaseolus  15 FTLPIGIQV   IPLFCKCPSK  NQLDRGIKYL  ITHVWQPNDN  VSFVSNKLGA   200

                       21         22         23         24         250
Lotus      20 SPADILTENR  YGODFTAATN  LPILIPVTQ   PELTQPSSNG  RKSSIHLLV    250
Glycine    20 SPEDIMSENN  YGQNFTAANN  LPVLIPVTR   PVLARSPSDG  RKGGIRLPVI   250
Phaseolus  20 SPQDILSENN  YGQNFTAASN  LPVLIPVTL   PDLIQSPSDG  RKHRIGLPVI   250

                       26         27         28         29         300
Lotus      25 LGITLGCTL   TAVLTGTLVY  VYCRRKKALN  RTASSAETAD  KLLSGVSGYV   300
Glycine    25 IGISLGCTL   VLVLAVLLVY  VYCLKMKTLN  RSASSAETAD  KLLSGVSGYV   300
Phaseolus  25 IGISLGCTL   VVVSAILLVC  VCCLKMKSLN  RSASSAETAD  KLLSGVSGYV   300

                       31         32         33         34         350
Lotus      30 SKPNVYEIDE  IMEATKDFSD  ECKVGESVYK  ANIEGRVVAV  KKIKEGGANE   350
Glycine    30 SKPTMYETDA  IMEATMNLSE  QCKIGESVYK  ANIEGKVLAV  KRFKED-VTE   350
Phaseolus  30 SKPTMYETGA  ILEATMNLSE  QCKIGESVYK  ANIEGKVLAV  KRFKED-VTE   350

                       36         37         38         39         400
Lotus      35 ELKILQKVNH  GNLVKLMGVS  SGYDGNCFLV  YEYAENGSLA  BWLFSKS--    400
Glycine    35 ELKILQKVNH  GNLVKLMGVS  SDNDGNCFVV  YEYAENGSLD  EWLFSKSCSD   400
Phaseolus  35 ELKILQKVNH  GNLVKLMGVS  SDNDGNCFVV  YEYAENGSLE  EWLFAKSCSE   400

                       41         42         43         44         450
Lotus      40 -SGTPNSLTW  SQRISIAVDV  AVGLQYMHEH  TYPRIIHRD   TTSNILLDSN   450
Glycine    40 TSNSRASLTW  CQRISMAVDV  AMGLQYMHEH  AYPRIVHRDI  TSSNILLDSN   450
Phaseolus  40 TSNSRTSLTW  CQRISIAVDV  SMGLQYMHEH  AYPRIVHRDI  TSSNILLDSN   450

                       46         47         48         49         500
Lotus      45 FKAKIANFAM  ARTSTNPMMP  KIDVFAFGVL  LIELLTGRKA  MTTKENGEVV   500
Glycine    45 FKAKIANFSM  ARTFTNPMMP  KIDVFAFGVV  LIELLTGRKA  MTTKENGEVV   500
Phaseolus  45 FKAKIANFSM  ARTFTNPMMS  KIDVFAFGVV  LIELLTGRKA  MTTKENGEVV   500

                       51         52         53         54         550
Lotus      50 MLWKDMWEIF  DIEENREERI  RKWMDPNLES  FYHIDNALSL  ASLAVNCTAD   550
Glycine    50 MLWKDIWKIF  DQEENREERL  KKWMDPKLES  YYPIDYALSL  ASLAVNCTAD   550
Phaseolus  50 MLWKDIWKIF  DQEENREERL  RKWMDPKLDN  YYPIDYALSL  ASLAVNCTAD   550

                       56         57         58         59         600
Lotus      55 KSLSRPSMAE  IVLSLSFLT   QSSNPTLERS  LTSSGLDVED  DAHIVTSIT    600
Glycine    55 KSLSRPTIAE  IVLSLSLLT   PSP-ATLERS  LTSSGLDVEA  -            600
Phaseolus  55 KSLSRPTIAE  IVLSLSLLT   PSP-ATLERS  LTSSGLDVEA  -            600

                       61         62         63         64         650
Lotus      60 R....       ....        ....        ....        ....         650
Glycine    60 R....       ....        ....        ....        ....         650
Phaseolus  60 R....       ....        ....        ....        ....         650
```

Table 2

A. Sequence identity (%) between NFR5 cDNA coding sequences determined by pairwise sequence comparisons using NCBI BlastN

	Lj	Pv	Gm
Lj	100		
Pv	86	100	
Gm	80	90	100

B. Sequence identity (%) between NFR5 protein sequences determined by pairwise sequence comparisons NCBI BlastP

	Lj	Pv	Gm
Lj	100		
Pv	70	100	
Gm	73	86	100

Lj=*Lotus japonicus*, Pv=*Phaseolus vulgaris*, Gm=*Glycine max*

Table 3

Alignment of *Lotus* and *Pisum* NFR1 protein sequences

```
             1          2          3          4          50
Pisum    1   MKLKNGLLLF  F-.....    KVESKCVIGC DIALASYYVM P-.....     50
Pisum    1   MKLKNGLLLF  F-.....    KVDSKCVKGC DLALASYYVM P-.....     50
Lotus    1   MKLKTGLLLF  FILLLGHVC  HVESNCLKGC DLALASYYI  PGVFILQNI   50

             6          7          8          9          100
Pisum    5   TFMQSKLVTN  SFEVIVRYNR DIVFSNDNLF SYFRVNIPFP CECIGGEFLG  100
Pisum    5   NYMQSKIVTN  SSDVLNSYNK VLVTNHGNIF SYFRINIPF  CECIGGEFLG  100
Lotus    5   TFMQSEIVSS  N-.....    DKILNDINI  SFQRLNIPFP CDCIGGEFLG  100

             11         12         13         14         150
Pisum    10  HVFEYTANEG  DTYDLIANTY YASLTTVEVL KKYNSYDPNH IPVKAKVNVT  150
Pisum    10  HVFEYTTKKG  DTYDLIANNY YVSLTSVELL KFNSYDPNH  IPAKAKVNVT  150
Lotus    10  HVFEYSASKG  DTYETIANL  YANLTTVDLL KRFNSYDPKN IPVNAKVNVT  150

             16         17         18         19         200
Pisum    15  VNCSCGNSQI  SKDYGLFITY PLRPRDTLBK IARHSNLDEG VIQSYNLGVN  200
Pisum    15  VNCSCGNSQI  SKDYGLFVTY PLRSTDSLEK IANESKLDEG LIQNFNPDVN  200
Lotus    15  VNCSCGNSQV  SKDYGLFITY PIRPGDTLQD IANQSSLDAG LIQSFNPSVN  200

             21         22         23         24         250
Pisum    20  FSKGSGVVFP  PGRDKNGEYV PLYPRT-GLG KGAAAGISI  GIFALLLF    250
Pisum    20  FSRGSGIVF   PGRDKNGEYV PLYPKT-GVG KGVAIGISI  GVFAVLLFV   250
Lotus    20  FSKDSGIAF   PGRYKNGVYV PLYHRTAGLA SGAAVGISI  GTFVLLLLA   250

             26         27         28         29         300
Pisum    25  CIYIKYFQK   EEEKTKLF-Q VSTALSAQD- -ASGSGEYET SGSSGHGTGS  300
Pisum    25  CIYVKYFQKK  EEEKTILF   VSKALSTQDG NASSSGEYET SGSSGHGTGS  300
Lotus    25  CMYVRY-QKK  EEEKAKLPTD ISMALSTQD- -ASSSAEYET SGSSGPTAS   300

             31         32         33         34         350
Pisum    30  TAGLTGIMVA  KSTEFSYQEL AKATNNFSLD NKIGGGFGA  VYYAVLRGEK  350
Pisum    30  AAGLTGIMVA  KSTEFSYQEL AKATDNFSLD NKIGQGGFGA VYYAELRGEK  350
Lotus    30  ATGLTSIMVA  KSMEFSYQEL AKATNNFSLD NKIGQGGFGA VYYAELRGKK  350

             36         37         38         39         400
Pisum    35  TAIKKMDVQA  STEFLCELQV LTHVHHLNLV RLIGYCVEGS LFLVYEHID   400
Pisum    35  TAIKKMNVQA  SSEFLCELKV LTHVHHLNLV RLIGYCVEGS LFLVYEHID   400
Lotus    35  TAIKKMDVQA  STEFLCELKV LTHVHHLNLV RLIGYCVEGS LFLVYEHID   400

             41         42         43         44         450
Pisum    40  GNLGQYLHGI  DKAPLPWSSR VQIALDSARG LEYIHEHTVP VYIHRDVKSA  450
Pisum    40  GNLGQYLHGK  DKEPLPWSSR VQIALDSARG LEYIHEHTVP VYIHRDVKSA  450
Lotus    40  GNLGQYLHGS  GKEPLPWSSR VQIALDAARG LEYIHEHTVP VYIHRDVKSA  450

             46         47         48         49         500
Pisum    45  NILIDKNLH   KVADFGLTKL IEVGNSTLHT RLVGTFGYMP PEYAQYGDVS  500
Pisum    45  NILIDKNLR   KVADFGLTKL IEVGNSTLHT RLVGTFGYMP PEYAQYGDVS  500
Lotus    45  NILIDKNLR   KVADFGLTKL IEVGNSTLQT RLVGTFGYMP PEYAQYGDIS  500

             51         52         53         54         550
Pisum    50  PKIDVYAFGV  VLYELISAK  AILKTGESAV ........   EEALNQIDPL  550
Pisum    50  PKIDVYAFGV  VLYELISAK  AVLKTGEESV AESKGLVALF EKALNQIDPS  550
Lotus    50  PKIDVYAFGV  VLFELISAK  AVLKTGE    AESKGLVALF EEALNKSDPC  550

             56         57         58         59         600
Pisum    55  EALRKLVDPR  LKENYPIDSV LKMAQLGRAC TRDNPLLRPS MRSLVVALMT  600
Pisum    55  EALRKLVDPR  LKENYPIDSV LKMAQLGRAC TRDNPLLRPS MRSLVVDLMT  600
Lotus    55  DALRKLVDPR  LGENYPIDSV LKIAQLGRAC TRDNPLLRPS MRSLVVALMT  600

             61         62         63         64         650
Pisum    60  LLSHTDD-    DTFYENQSLT NLLSVR     ....       ....        650
Pisum    60  LSSPFEDCDD  DTSYENQTLI NLLSVR     ....       ....        650
Lotus    60  LSSLTEDCDD  ESSYESQTLI NLLSVR     ....       ....        650
```

Table 4

A. **Sequence identity (%) between NFR1 cDNA coding sequences determined by pairwise sequence comparisons using NCBI BlastN**

	Lj	PsNFR1a	PsNFR1B
Lj	100		
PsNFR1A	84	100	
PsNFR1B	83	87	100

B. **Sequence identity (%) between NFR1 protein sequences determined by pairwise sequence comparisons NCBI BlastP**

	Lj	PsNFR1A	PsNFR1B
Lj	100		
PsNFR1A	73	100	
PsNFR1B	75	79	100

Lj=*Lotus japonicus*, Ps=*Pisum sativum*

Table 5

Summary of *Lotus nfr5* and pea *sym10* mutant alleles

Allele	Mutation	*Lotus* Spp
sym5-1	EYAENGSLA 380-388 deletion	Lj
sym5-2	retrotransposon integration after Q233	Lj
sym5-3	CAG→TAG, Q55→stop	Lj
RisFixG	TGG→TGA, W_{388}→stop	Ps
P5	TGG→TGA, W_{405}→stop	Ps
P56	CAA→TAA, Q_{200}→stop	Ps
N15	*Sym10* gene deleted	Ps

TABLE 6

Complementation of *Lotus japonicus nfr5* mutants with the wildtype *NFR5* transgene

Lotus genotype	Transgene	No. of plants	Infected With	No. of plants with nodules*	Total No. of nodules
nfr5-1	NFR5	31	*M.loti* NZP2235	18	nd
nfr5-1	Empty vector	20	*M.loti* NZP2235	0	nd
nfr5-2	NFR5	5	*M.loti* NZP2235	1	nd
nfr5-2	Empty vector	5	*M.loti* NZP2235	0	nd

* Nodules only detected on transformed roots

TABLE 7

Transformation of *Lotus japonicus nfr1* mutants with the wildtype *NFR1* transgene

Lotus genotype	Transgene	No. of plants	Infected With	No. plants with nodules	Total No. of nodules	Average No. nodules/ plant
nfr1-1	NFR1	103	*M.loti* NZP2235	62*	310	5
nfr1-1	Empty vector	30	*M.loti* NZP2235	0	0	0
nfr1-2	NFR1	20	*M.loti* NZP2235	13*	97	7.5
nfr1-2	empty vector	7	*M.loti* NZP2235	0	0	0

* Nodules only detected on transformed roots

Table 8

Lotus filicaulis transformed with wildtype *NFR1* and *NFR5* genes from *Lotus japonicus* Gifu

Lotus genotype	Transgene	No. of plants	Infected with	No. plants with nodules	Total No. of nodules	Average No. nodules/ plant
Lotus filicaulis	NFR1+ NFR5	104	R.leg 5560 DZL	10*	25	2.5
Lotus filicaulis	Empty vector	65	R.leg 5560 DZL	0	0	0
Lotus japonicus Gifu	Empty vector	10	R.leg 5560 DZL	10**	>150	>15

* Nodules only detected on transformed roots
** Nodules on normal and transformed roots

Table 9

L. japonicus nfr5 mutant transformed with a hybrid _NFR5_ gene "FinG5" encoding the extracellular domain of _L.filicaulis_ NFR5 fused to the kinase domain from _L. japonicus_ Gifu NFR5.

Lotus genotype	Transgene	No. of plants	Infected with	No. of plants with nodules	Total No. of nodules	Average No. nodules/ plant
nfr5	FinG5	31	M.loti NZP2235	28*	~180	6.4
nfr5	Empty vector	12	M.loti NZP2235	0	0	
nfr5	FinG5	34	R.leg 5560 DZL	1*	4	4 1 PLANT ONLY
nfr5	empty vector	10	R.leg 5560 DZL	0	0	
Lotus japonicus Gifu	empty vector	10	R.leg 5560 DZL	10**	>150	>15
Lotus filicaulis	empty vector	29	R.leg 5560 DZL	0	0	

* Nodules only detected on transformed roots
** Nodules on normal and transformed roots

Table 10

L. japonicus nfr1 mutant transformed with a hybrid NFR1 gene "FinG1" encoding the extracellular domain of L.filicaulis NFR1 fused to the kinase domain from L. japonicus Gifu NFR1.

Lotus genotype	Transgene	No. of plants	Infected with	No. of plants with nodules	Total No. of nodules	Average No. nodules/ plant
nfr1-1	FinG1	8	M.loti NZP2235	8*	59	7.3
nfr1-1	Empty vector	6	M.loti NZP2235	0	0	0
nfr1-1	FinG1	13	R. leg 5560DZL	5*#	15	3
nfr1-1	Empty vector	9	R. leg 5560DZL	0	0	0
nfr1-2	FinG1	10	R. leg 5560DZL	3*#	12	4
nfr1-2	Empty vector	4	R. leg 5560DZL	0	0	0

* Nodules only detected on transformed roots
# Nodules were first counted after 56 days, while M.loti NZP2235 nodules were detectable after ~25 days.

Table 11

Molecular markers for *NFR5* allele breeding in *Lotus*

Marker	Genetic distance from *NFR5* locus	*Lotus* Ecotype	Microsatellite sequence
TM0272	2,9cM	MG-20	18xCT
		Gifu	12xCT
TM0257	1,0cM	MG-20	10xAAG
		Gifu	7xAAG
LjT13i23Sfi		Gifu	TTTTGCTGCAGCAAGTCAGACTGTTAGAGGA
		Fili	TTTTGCTGCAACAAGTCGGACTGTTAGAGGA
TM0522	0cM	MG-20	24xAT
		Gifu	14xAT
NFR5			
E32M54-12F	0,5cM	MG-20	TTGGAAGTTCTTTTTA**TTA**GGTTAATTTTA
		Fili	TTGGAAGTTCTTTTTA---GGTTAATTTTA
LjT01c03 Not	0,7cM	Fili	CATTCCAGAAGAAAATAAGATATAATTATG
		MG-20	CATTCCAGAAGAAAATAAGATATAATTATG
		Gifu	CATTCCAGAAG-AAATAAGATATAATTATG
TM0168	2,2cM	MG-20	19xAT
		Gifu	15xAT
TM0021	3,8cM	MG-20	16xCT
		Gifu	13xCT

Table 12

Nucleotide sequence variation between the pea *SYM10* alleles
of pea cultivars Frisson and Finale*

```
Frisson    CTTGCATTTC TTCACAATTT CACAACAATG GCTATCTTCT TTCTTCCTTC
Finale     CTTGCATTTC TTCACAATTT CACAACAATG GCTATCTTCT TTCTTCCTTC

Frisson    TAGTTCTCAT GCCCTTTTTC TTGCACTCAT GTTTTTTGTC ACTAATATTT
Finale     TAGTTCTCAT GCCCTTTTTC TTGCACTCAT GTTTTTTGTC ACTAATATTT

Frisson    CAGCTCAACC ATTACAACTC AGTGGAACAA ACTTTTCATG CCCGGTGGAT
Finale     CAGCTCAACC ATTACAACTC AGTGGAACAA ACTTTTCATG CCCGGTGGAT

Frisson    TCACCTCCTT CATGTGAAAC CTATGTGACA TACTTTGCTC GGTCTCCAAA
Finale     TCACCTCCTT CATGTGAAAC CTATGTGACA TACTTTGCTC GGTCTCCAAA

Frisson    CTTTTTGAGC CTAACTAACA TATCAGATAT ATTTGATATG AGTCCTTTAT
Finale     CTTTTTGAGC CTAACTAACA TATCAGATAT ATTTGATATG AGTCCTTTAT

Frisson    CCATTGCAAA AGCCAGTAAC ATAGAAGATG AGGACAAGAA GCTGGTTGAA
Finale     CCATTGCAAA AGCCAGTAAC ATAGAAGATG AGGACAAGAA GCTGGTTGAA

Frisson    GGCCAAGTCT TACTCATACC TGTAACTTGT GGTTGCACTA GAAATCGCTA
Finale     GGCCAAGTCT TACTCATACC TGTAACTTGT GGTTGCACTA GAAATCGCTA

Frisson    TTTCGCGAAT TTCACGTACA CAATCAAGCT AGGTGACAAC TATTTCATAG
Finale     TTTCGCGAAT TTCACGTACA CAATCAAGCT AGGTGACAAC TATTTCATAG

Frisson    TTTCAACCAC TTCATACCAG AATCTTACAA ATTATGTGGA AATGGAAAAT
Finale     TTTCAACCAC TTCATACCAG AATCTTACAA ATTATGTGGA AATGGAAAAT

Frisson    TTCAACCCTA ATCTAAGTCC AAATCTATTG CCACCAGAAA TCAAAGTTGT
Finale     TTCAACCCTA ATCTAAGTCC AAATCTATTG CCACCAGAAA TCAAAGTTGT

Frisson    TGTCCCTTTA TTCTGCAAAT GCCCCTCGAA GAATCAGTTG AGCAAAGGAA
Finale     TGTCCCTTTA TTCTGCAAAT GCCCCTCGAA GAATCAGTTG AGCAAAGGAA

Frisson    TAAAGCATCT GATTACTTAT GTGTGGCAGG CTAATGACAA TGTTACCCGT
Finale     TAAAGCATCT GATTACTTAT GTGTGGCAGG CTAATGACAA TGTTACCCGT

Frisson    GTAAGTTCCA AGTTTGGTGC ATCACAAGTG GATATGTTTA CTGAAAACAA
Finale     GTAAGTTCCA AGTTTGGTGC ATCACAAGTG GATATGTTTA CTGAAAACAA

Frisson    TCAAAACTTC ACTGCTTCAA CCAACGTTCC GATTTTGATC CCTGTGACAA
Finale     TCAAAACTTC ACTGCTTCAA CCAATGTTCC GATTTTGATC CCTGTGACAA

Frisson    AGTTACCGGT AATTGATCAA CCATCTTCAA ATGGAAGAAA AAACAGCACT
Finale     AGTTACCGGT AATTGATCAA CCATCTTCAA ATGGAAGAAA AAACAGCACT
```

```
Frisson    CAAAAACCTG CTTTTATAAT TGGTATTAGC CTAGGATGTG CTTTTTTCGT
Finale     CAAAAACCTG CTTTTATAAT TGGTATTAGC CTAGGATGTG CTTTTTTCGT

Frisson    TGTAGTTTTA ACACTATCAC TTGTTTATGT ATATTGTCTG AAAATGAAGA
Finale     TGTAGTTTTA ACACTATCAC TTGTTTATGT ATATTGTCTG AAAATGAAGA

Frisson    GATTGAATAG GAGTACTTCA TTGGCGGAGA CTGCGGATAA GTTACTTTCA
Finale     GATTGAATAG GAGTACTTCA TTGGCGGAGA CTGCGGATAA GTTACTTTCA

Frisson    GGTGTTTCGG GTTATGTAAG CAAGCCAACA ATGTATGAAA TGGATGCGAT
Finale     GGTGTTTCGG GTTATGTAAG CAAGCCAACA ATGTATGAAA TGGATGCGAT

Frisson    CATGGAAGCT ACAATGAACC TGAGTGAGAA TTGTAAGATT GGTGAATCCG
Finale     CATGGAAGCT ACAATGAACC TGAGTGAGAA TTGTAAGATT GGTGAATCTG

Frisson    TTTACAAGGC TAATATAGAT GGTAGAGTTT TAGCAGTGAA AAAAATCAAG
Finale     TTTACAAGGC TAATATAGAT GGTAGAGTTT TAGCAGTGAA AAAAATCAAG

Frisson    AAAGATGCTT CTGAGGAGCT GAAAATTTTG CAGAAGGTAA ATCATGGAAA
Finale     AAAGATGCTT CTGAGGAGCT GAAAATTCTG CAGAAGGTAA ATCATGGAAA

Frisson    TCTTGTGAAA CTTATGGGTG TGTCTTCCGA CAACGACGGA AACTGTTTCC
Finale     TCTTGTGAAA CTTATGGGTG TGTCTTCCGA CAACGAAGGA AACTGTTTCC

Frisson    TTGTTTACGA GTATGCTGAA AATGGATCAC TTGATGAGTG GTTGTTCTCA
Finale     TTGTTTACGA GTATGCTGAA AATGGATCAC TTGATGAGTG GTTGTTCTCA

Frisson    GAGTCGTCGA AAACTTCGAA CTCGGTGGTC TCGCTTACAT GGTCTCAGAG
Finale     GAGTTGTCGA AAACTTCGAA CTCGGTGGTC TCGCTTACAT GGTCTCAGAG

Frisson    AATAACAGTA GCAGTGGATG TTGCAGTTGG TTTGCAATAC ATGCATGAAC
Finale     AATAACAGTA GCAGTGGATG TTGCAGTTGG TTTGCAATAC ATGCATGAAC

Frisson    ATACTTACCC AAGAATAATC CACAGAGACA TCACAACAAG TAATATCCTT
Finale     ATACTTACCC AAGAATAATC CACAGAGACA TCACAACAAG TAATATCCTT

Frisson    CTGGATTCAA ACTTTAAGGC CAAGATAGCG AATTTTTCAA TGGCCAGAAC
Finale     CTGGATTCAA ACTTTAAGGC CAAGATAGCG AATTTTTCAA TGGCCAGAAC

Frisson    TTCAACAAAT TCCATGATGC CGAAAATCGA TGTTTTCGCT TTTGGGGTGG
Finale     TTCAACAAAT TCCATGATGC CGAAAATCGA TGTTTTCGCT TTTGGGGTGG

Frisson    TTCTGATTGA GTTGCTTACC GGCAAGAAAG CGATAACAAC GATGGAAAAT
Finale     TTCTGATTGA GTTGCTTACC GGCAAGAAAG CGATAACAAC GATGGAAAAT

Frisson    GGCGAGGTGG TTATTCTGTG GAAGGATTTC TGGAAGATTT TTGATCTAGA
Finale     GGCGAGGTGG TTATTCTGTG GAAGGATTTC TGGAAGATTT TTGATCTAGA

Frisson    AGGGAATAGA GAAGAGAGCT TAAGAAAATG GATGGATCCT AAGCTAGAGA
Finale     AGGGAATAGA GAAGAGAGCT TAAGAAAATG GATGGATCCT AAGCTAGAGA
```

```
Frisson    ATTTTTATCC TATTGATAAT GCTCTTAGTT TGGCTTCTTT GGCAGTGAAT
Finale     ATTTTTATCC TATTGATAAT GCTCTTAGTT TGGCTTCTTT GGCAGTGAAT

Frisson    TGTACTGCAG ATAAATCATT GTCAAGACCA AGCATTGCAG AAATTGTTCT
Finale     TGTACTGCAG ATAAATCATT GTCAAGACCA AGCATTGCAG AAATTGTTCT

Frisson    TTGTCTTTCT CTTCTCAATC AATCATCATC TGAACCAATG TTAGAAAGAT
Finale     TTGTCTTTCT CTTCTCAATC AATCATCATC TGAACCAATG TTAGAAAGAT


Frisson    CCTTGACATC TGGTTTAGAT GTTGAAGCTA CTCATGTTGT TACTTCTATA
Finale     CCTTGACATC TGGTTTAGAT GTTGAAGCTA CTCATGTTGT TACTTCTATA

Frisson    GTAGCTCGTT GATATTCATT CAAGTGAAGG TAACACTGAA TCAATGCTTC
Finale     GTAGCTCGTT GATATTCATT CAAGTGAAGG TAACACTAAA TCAATGCTTC

Frisson    AGTTTCTTAT ATTCAAGATG GTTACTTTGT TTAGATGATT ATTGATTACA
Finale     AGTTTCTTAT ATTCAAGATG GTTACTTTGT TTAGGTGATT ATTGATTACA

Frisson    TCTTTATGTG TGGAACTATA TGGTTATTTT AATTAAGGGA ATTGTTCTAA
Finale     TCTTTATGTG TGGAACTATA TGGTTATTTT AATTAAGGGA ATTAGTCTAA

Frisson    AATTCATTTT TCCATGTT
Finale     ATTTCATTTT TCCATGTT
```

* Nucleotide differences are shaded black and the coding region is underlined

Table 13

Protein sequence differences encoded by the pea *SYM10* alleles of pea cultivars Frisson and Finale*

Frisson	MAIFFLPSSS	HALFLALMFF	VTNISAQPLQ	LSGTNFSCPV	DSPPSCETYV
Finale	MAIFFLPSSS	HALFLALMFF	VTNISAQPLQ	LSGTNFSCPV	DSPPSCETYV
Frisson	TYFARSPNFL	SLTNISDIFD	MSPLSIAKAS	NIEDEDKKLV	EGQVLLIPVT
Finale	TYFARSPNFL	SLTNISDIFD	MSPLSIAKAS	NIEDEDKKLV	EGQVLLIPVT
Frisson	CGCTRNRYFA	NFTYTIKLGD	NYFIVSTTSY	QNLTNYVEME	NFNPNLSPNL
Finale	CGCTRNRYFA	NFTYTIKLGD	NYFIVSTTSY	QNLTNYVEME	NFNPNLSPNL
Frisson	LPPEIKVVVP	LFCKCPSKNQ	LSKGIKHLIT	YVWQANDNVT	RVSSKFGASQ
Finale	LPPEIKVVVP	LFCKCPSKNQ	LSKGIKHLIT	YVWQANDNVT	RVSSKFGASQ
Frisson	VDMFTENNQN	FTASTNVPIL	IPVTKLPVID	QPSSNGRKNS	TQKPAFIIGI
Finale	VDMFTENNQN	FTASTNVPIL	IPVTKLPVID	QPSSNGRKNS	TQKPAFIIGI
Frisson	SLGCAFFVVV	LTLSLVYVYC	LKMKRLNRST	SLAETADKLL	SGVSGYVSKP
Finale	SLGCAFFVVV	LTLSLVYVYC	LKMKRLNRST	SLAETADKLL	SGVSGYVSKP
Frisson	TMYEMDAIME	ATMNLSENCK	IGESVYKANI	DGRVLAVKKI	KKDASEELKI
Finale	TMYEMDAIME	ATMNLSENCK	IGESVYKANI	DGRVLAVKKI	KKDASEELKI
Frisson	LQKVNHGNLV	KLMGVSSDN**D**	GNCFLVYEYA	ENGSLDEWLF	SE**S**SKTSNSV
Finale	LQKVNHGNLV	KLMGVSSDN**E**	GNCFLVYEYA	ENGSLDEWLF	SE**L**SKTSNSV
Frisson	VSLTWSQRIT	VAVDVAVGLQ	YMHEHTYPRI	IHRDITTSNI	LLDSNFKAKI
Finale	VSLTWSQRIT	VAVDVAVGLQ	YMHEHTYPRI	IHRDITTSNI	LLDSNFKAKI
Frisson	ANFSMARTST	NSMMPKIDVF	AFGVVLIELL	TGKKAITTME	NGEVVILWKD
Finale	ANFSMARTST	NSMMPKIDVF	AFGVVLIELL	TGKKAITTME	NGEVVILWKD
Frisson	FWKIFDLEGN	REESLRKWMD	PKLENFYPID	NALSLASLAV	NCTADKSLSR
Finale	FWKIFDLEGN	REESLRKWMD	PKLENFYPID	NALSLASLAV	NCTADKSLSR
Frisson	PSIAEIVLCL	SLLNQSSSEP	MLERSLTSGL	DVEATHVVTS	IVAR
Finale	PSIAEIVLCL	SLLNQSSSEP	MLERSLTSGL	DVEATHVVTS	IVAR

* Amino acid differences are highlighted in black.